ANGLIA RUSKIN UNIVERSITY

ASSESSING SERVICE PERFORMANCE: APPLYING DATA ENVELOPMENT ANALYSIS FOR EVALUATING EMPLOYEES´ PERFORMANCE IN THE SERVICE INDUSTRY

by

MANUELA KOCH-ROGGE

A thesis in partial fulfilment of the requirements of Anglia Ruskin University for the degree of Doctor of Philosophy

Submitted: April 2019

Acknowledgements

Taking the first steps into the academic community is a unique experience. Working on a thesis is a challenge that spans over years involving amounts of literature to read, theories to understand, different opinions to evaluate, studies to carry out and conclusions to draw. Along this journey, there are many exiting and enlightening moments, like presenting parts of your work to academic colleagues for the first time or getting your paper accepted by scientific journals. However, there are also many obstacles and moments of frustration along the way. For me, this includes following a particular direction of research that leads to a dead end or just fighting with editing and formatting issues. It´s in particular during those times that some people that follow you on that journey become most important.

First of all, I would like to thank my supervisors, who supported me in complementary ways. I thank Professor Georg Westermann, who facilitated my dissertation at Anglia Ruskin University in the first place and whose support was the main guidance towards the completion of this thesis. I also thank Dr Chris Wilbert, Prof Georg Westermann and Dr Rob Willis for their valuable comments and expertise when conducting my thesis, but also for their reassurance to present my work to the academic community. Also, I am very grateful for the financial support of ARU´s Lord Ashcroft International Business School.

I thank all the people who took part in my empirical work, in particular the ones who agreed to participate in the case study and the subsequent assessment. My thanks also go to my department colleagues and the network of researchers who I worked with and met during my research.

Very special thanks go to my family. I thank my husband, Steffen, who not only supported me emotionally but also professionally. Finally, I would like to thank my parents, Martina and Bernd, for their constant support and encouragement during these years.

Thesis abstract

The service industry is the most important economic sector worldwide. Thus, economic growth within developed economies is almost exclusively based on services. In the production of services, employees have a crucial influence on the perceived quality of the output, since they are the core of the interaction with the customers. In this context, employees´ knowledge and skills are primary resources for an organization´s ability to compete and generate profits. The adequate evaluation of employee performance, employee empowerment and a concern for training therefore becomes a necessity for each service organization. However, traditional approaches for evaluating employee performance mostly originate from the manufacturing sector and therefore often fall short in accounting for the advanced requirements of evaluating service performance. A promising approach that may be able to address several of the shortcomings of traditional performance evaluation approaches is the Data Envelopment Analysis (DEA), which has widely been used to evaluate the performance of organizations in a variety of contexts. However, there is still very little experience regarding the application of DEA on individual level and employee´s response to it.

Adopting a case-study strategy based on a mixed methods approach, this research investigates DEA´s technical and organisational suitability for evaluating employees´ performance in the service sector. After reviewing literature on why services may require an advanced approach for performance evaluation and on DEA´s previous applications in organizational contexts, the DEA approach was applied to performance data of 40 service employees in a German Cooperative Bank to gain empirical evidence. Hence, in a quantitative analysis, the results of the application were examined to assess DEA´s technical suitability and to gain a deeper understanding of its application on individual level. Subsequently, focus group interviews among bank managers and members of the workers´ council as well as a questionnaire study among employees were carried out to investigate DEA´s perception by all major stakeholders and thus provide substantiation of its organizational suitability. Eventually, the evidence from these multiple sources was analysed by triangulation.

The findings indicate that DEA is a suitable approach to overcome many shortcomings of traditional approaches. Thus, the thesis concludes by summarizing the results, reflecting on limitations and by pointing out implications that may contribute to draw analytic generalizations to inform theory and practice.

.

Table of Contents

List of figures

List of tables

1 Introduction

"The bad news is that ignoring the performance of people is almost as bad as shredding their effort in front of their eyes. (…) The good news is that by simply looking at something that somebody has done, scanning it and saying "uh huh," [you] dramatically improves people's motivations."

Dan Ariely, Professor of Psychology and Behavioural Economics at Duke University, US (Ariely, 2013)

" Any business or industry that pays equal rewards to its goof-offs and eager beavers sooner or later will find itself with more goof-offs than eager beavers."

Mick Delaney, American college football coach and former head coach at the University of Montana, US (Yeates, 2015)

"It should be your priority looking after your number one customer first: your employee."

Ian Hutchinson, Chief Engagement Officer and founder of Life by Design® (Hutchinson, 2009)

Looking at the quotes above, it becomes evident that the recognition and evaluation of individual performance touch everybody´s life to some extent. It is subject to research (Ariely, 2013) a basis for sporting successes (Delaney in (Yeates, 2015)) and of course a main task in Human Resources Management (Hutchinson, 2009). Also, individual performance evaluation has many different functions including employee motivation, the distribution of rewards or identifying the key drivers of organizational success. Even though it a necessary and important task, evaluating employees´ performance is often regarded with ambiguity by both researchers and practitioners. Particularly in the service industry, the adequate evaluation of individual performance faces many obstacles.

This doctoral thesis sheds light on the importance of evaluating employee performance in the service industry adequately and investigates the requirements to an adequate performance evaluation approach on individual level. Subsequently, the thesis illustrates the development of a performance evaluation approach based on Data Envelopment Analysis and further provides a comprehensive assessment involving all relevant stakeholders.

This introductory chapter illustrates the research background and the research problem by outlining the role of employees in the service industry and pointing out the issues involved when evaluating employees´ performance. Subsequently, Data Envelopment Analysis (DEA) is introduced as a potential approach to meet those requirements. Based on this, the research aim of this thesis is illustrated, the research questions are derived and the delimitation of this research is discussed. Finally, the chapter provides an overview of the research methodology used and concludes by illustrating the structure of the thesis.

1.1 The Role of Employee Performance in the Service Industry

The service sector is expanding rapidly worldwide and service markets have never been larger. Hence, the economic growth within developed countries is almost exclusively based on services (Lovelock, Gummesson 2004, Fitzsimmons, Fitzsimmons 2008) with business services like finance, renting or R&D being the key drivers (Wirtz and Ehret, 2017). In Germany, service industries generated about 70 per cent of economic value in 2015 and currently contribute up to 75 per cent to employment (German Federal Ministry for Economic Affairs and Energy 2015). However, in spite of its crucial importance, the research into productivity and efficiency of services is still scarce (Grönroos and Ojasalo, 2004; C. Lovelock and Gummesson, 2004) and there are yet no agreed answers concerning the “productivity gap” in the service sector (Benkenstein et al., 2017). Due to the intangible and heterogeneous nature of services and the inseparability of production and consumption, traditional productivity concepts are too limited for evaluating service productivity and the underlying assumptions of these concepts often do not hold. For instance, by applying traditional productivity concepts on services, only measurements of partial productivity are obtained, since multiple production factors cannot be integrated. At the same time, the effects of productivity improvement on the economic results or on customer value cannot be controlled. Thus, “improved” productivity may negatively affect perceived service quality, customer value and finally economic results (Grönroos and Ojasalo, 2004).

Since service employees have a crucial influence on the perceived quality of the output, they are among the most important input resources in the production of services (Vargo, Lusch 2008, Fitzsimmons, Fitzsimmons 2008, Farquhar 2004). They interact with customers during the service process and often build long-term relationships with them. In turn, the productivity of services is frequently dependent on how these relationships between service provider and customer develop (Benkenstein et al., 2017). Several studies point out the significance of perceived service quality and customer satisfaction

to customer loyalty and retention (Hallowell 1996, Farquhar 2004). In this context, employees´ knowledge and skills are primary resources that are often undervalued in the process of value-creation and customer relationship (Grönroos and Voima, 2013). Thus, service employees often need to be highly qualified with skills that are difficult to acquire. In this respect, adequate assessment of employee´s performance, employee empowerment and a concern for training become increasingly necessary for continuing competitiveness (Fitzsimmons, Fitzsimmons 2008).

On the other hand, employees´ performance is often affected by an intertwining of personal and environmental factors like employee´s experience, education or social competencies as well as by varying levels of customer participation in the service process. Regarding these issues, the prerequisites and framework conditions to render a specified service or to achieve a performance target may vary from one employee to another. Therefore, it seems reasonable that employees use different input-output-processes in order to produce a similar service. Accordingly, the suggestion of a parametric identical production function seems unsuitable for services. Furthermore, some of the input and output factors to produce a service are even beyond the employee´s control (e.g. competitive environment). Based on this assumption, only employees with similar production processes who face similar environmental conditions should be compared to one another or should serve as peers to others. In order to provide feedback and to point out potential for improving competencies, a performance evaluation should also ensure to capture the employee´s total performance and not only partial aspects (Kline, Sulsky 2009).

When it comes to assess and measure service employee´s performance, the underlying evaluation method should account for all those requirements. Thus, The inadequacy of extant approaches for assessing service performance and with this the choice of an adequate performance evaluation method is a recurring, but unsolved issue in service sector research (Boles et al., 1995; Ostrom et al., 2015).

1.2 Requirements for evaluating Employee Performance in the Service Sector

The evaluation of employee performance is at the core of Human Resources Management activities. However, role of Human Resources Management (HRM) including performance evaluation faced some major criticism during the last decade, considering the question how HRM removed itself from its original concern for the well-being of individuals (Ardichivili, 2013) and may have contributed to the recent global financial crisis (GFC) (MacKenzie et al., 2012).

Moreover, many managers show discomfort towards evaluating the performance of their subordinates (C. E. Pettijohn et al., 2001), some even describe them as onerous and distasteful (Yen et al., 2017). To a great extent, this may be due to the requirements to a performance evaluation on the one hand, and the rather poor range of evaluation methods that are able to account for those requirements on the other hand. In Performance Management research, considerable effort has been carried out during the last decades to determine what constitutes a good performance evaluation. However, due to the mentioned characteristics of services, traditional performance evaluation methods have not been very effective in identifying superior performance and disseminating best practice in service organizations (Schaffnit et al., 1997). Neely et al. (2000) claim that most currently applied performance evaluation methods or frameworks are adopted from the manufacturing sector and therefore have a narrow, often unidimensional focus.

The majority of studies highlight the fact that the performance evaluation needs to be accepted and perceived as fair by the employees to improve performance or increase job satisfaction (Greenberg 1986, Meyer 1991, Cocca, Alberti 2010, Sudin 2011). Otherwise, performance is actually likely to decrease (Flint 1999). But what makes a performance evaluation a fair one? One of the most cited is rater´s bias (Kline and Sulsky, 2009; Kondrasuk, 2012), which may be due to subjective perception or even rating errors like the so called “Halo effect”. Another issue that relates to perceived fairness is “external” or “environmental” effects that may affect employees´ performance, but which are beyond their control (Paradi and Zhu, 2013). Especially in the service industry, environmental factors like opening hours, economic growth rate or competitive environment play a major role, since they may influence both the production and the perceived quality of a service. Thus, an adequate performance evaluation method needs to ensure non-biased results and, at the same time, should be able to control environmental effects. Again, most performance evaluation methods that are currently applied in practice fail to account for those requirements on employee level.

1.3 DEA for evaluating employee performance in the Service Sector

A promising approach that may be able to address several of the shortcomings of traditional performance evaluation approaches is the Data Envelopment Analysis (DEA), which was first introduced by Charnes, Cooper and Rhodes in 1978. The DEA methodology evaluates the performance of Decision Making Units (DMU) relative to all other units of a sample and is able to handle multiple input- and output figures simultaneously in complex operating situations. By transforming all inputs and outputs into a scalar measure, DEA computes a single score for each DMU. Since DEA is of non-parametric nature, it requires no a priori specifications of the parametric form of the production correspondences (Banker, Morey 1986a). Thus, it accounts for parametrically different production processes. Due to its non-parametric nature, DEA computes the weights for the inputs and outputs of each DMU on an individual basis. Although this liberalism does not necessarily lead to an efficiency of 100 per cent this principle allows each DMU to "shine in its best light" (Coughlan et al., 2010; Doyle and Green, 1994).

A DMU is deemed to be efficient (score of $\theta = 1$) if its output is optimal for its inputs in comparison to of inputs and outputs of all other DMUs under consideration. In this case the DMU cannot increase its output without increasing one or more inputs or decreasing another output. This is often referred to as "Pareto Optimality" (Donthu, Yoo 1998). The DMUs that exhibit best practice ($\theta = 1$) are identified and form an efficient frontier. Inefficient DMUs are enveloped by the frontier. For each DMU that is not on the frontier, DEA identifies a set of references that show similar input-output combinations, but are efficient. Thus, DEA develops a tailored comparison group for each inefficient DMU, whose members serve as peers and can help to identify the sources of inefficiency. DEA is also able to accommodate environmental factors (Moreno, Tadepalli 2002) and to incorporate qualitative data (Wagner et al., 2003).

Considering the complex nature of services and the requirements they place on the performance evaluation of service employees, DEA seems a suitable approach to meet the requirements outlined before. So far, DEA has been used to evaluate the performance of branch networks or organizations in a variety of contexts including hospitals (Du et al., 2014; Wagner et al., 2003) universities (B. L. Lee and Worthington, 2016; Nazarko and Šaparauskas, 2014; Tzeremes and Halkos, 2010) or airlines (Fethi et al., 2000; Wanke and Barros, 2016). However, despite its intense application on institutional and branch level, there is still very little experience regarding the application of DEA on individual level and even less experience considering its application in the service industry. In addition, there is a lack of information from both managers and employees considering adequateness or comprehensibility of the evaluation procedure

and results. In particular, literature holds no empirical data about employees´ response to a DEA based performance evaluation – neither on organizational nor on individual level.

1.4 The scope of the Thesis

Data Envelopment Analysis is a widely used method to assess the performance of different kinds of organizations and branches. At the same time, it has many qualities that seem suitable to address the recurring issues that arise when evaluating employee performance in the service sector. Hence, the aim of this thesis is to investigate whether DEA is an appropriate approach for evaluating performance on individual level. In the following, the research objectives and research questions of this thesis are posed. Furthermore, the delimitation of this research is outlined.

1.4.1 Research Objectives and Research Questions

The major aim of this thesis is to investigate whether DEA is a suitable approach for evaluating performance on individual level. Since "suitability" is a multi-layered, complex term, it should at least be considered from two distinct perspectives. On the one hand, there is a rather "technical" suitability that relates to the accuracy of results and to how the approach meets general requirements to a performance evaluation. However, being technically appropriate does not mean that an evaluation approach will be of practical use in HRM. Thus, another perspective to investigate is the "organizational" suitability of the proposed approach. This relates in particular to its appropriateness to contribute to organizational purposes and how it is received by the stakeholders concerned.

In this respect, this research aims to address the following research objectives and subsidiary research questions:

Objective A: Investigating DEA´s technical suitability for evaluating employees´ performance in the service sector.

This objective comprises the following subsidiary research questions:

A.1. Does the evaluation of employees´ performance in the Service Sector require an advanced method compared to the Production Sector?

A.2. What are the general requirements to a performance evaluation method? What are the requirements for an "advanced method"?

A.3. What are the pros and cons of existing performance evaluation methods?

A.4. Is DEA able to meet the requirements to an “advanced method” and to address shortcomings of traditional evaluation methods?

A.5. Does DEA provide accurate and meaningful results on individual level?

Objective B: Investigating DEA´s organizational suitability for evaluating employees´ performance in the service sector.

In consequence, the following subsidiary research questions arise:

B.1. How is the method perceived by the stakeholders concerned (in particular employees, management and workers´ council)?

B.2. Are the results suitable to base administrative and developmental decisions on?

B.3. Is the method suitable for determining individual performance targets or for calculating variable salary components (such as bonuses)?

1.4.2 Contribution of Research

The results of this research shall contribute to gather more experience of both theoretical and empirical nature on evaluating employees´ performance in the service industry. A particular focus is laid on the application of Data Envelopment Analysis, which has widely been applied on organizational and branch level, but with no comprehensive application on individual level in the service industry yet. Furthermore, there is a lack of empirical information about how non-traditional performance evaluation approaches are perceived by all parties involved. Whilst there are few studies about management´s response, there are no such data for employees or workers´ council. This thesis aims to address these shortfalls and to provide some insights into opportunities and limitations of evaluating service performance on individual level.

1.4.3 Delimitation

Charnes, Cooper and Rhodes presented the first DEA model, the so called CCR model, in 1978. This model was extended by Banker et al. (1984) to account for variable returns to scale. These two models are commonly referred to as “basic DEA models”. Since then, there has been a tremendous growth in both, theoretical development and applications to practical situations (Cook and Seiford, 2009). This research aims on applying DEA in a new context (measuring employee performance) by referring to the basic models. Thus, it shall contribute to examine DEA´s practicable applicability. Therefore, theoretical enhancements will only be discussed considering their applicability for the given purpose. Subsequently, this thesis is not aimed at contributing to the theoretical enhancement of DEA.

Performance Management is a broad topic that covers several interdisciplinary areas ranging from Human Resources Management, Organizational Management and Process Management to Strategic Management. Hence, it has been studied from varying perspectives including the organizational perspective and the individual perspective. Furthermore, Performance Management comprises numerous terms like Performance Measurement, Performance Evaluation or Performance Appraisal, many of them used simultaneously (Folan and Browne, 2005). However, this research aims to reflect the major issues of measuring employee performance. Therefore, the thesis first reflects on Performance Management in general, but then mainly focuses on the individual perspective. Furthermore, the term “Performance Evaluation” is used throughout the thesis for measuring and assessing performance.

For the practical application of the proposed approach, a case study in a German Cooperative bank was conducted. The case of banking services was selected due to several reasons. First, banking comprises a variety of different kinds of services ranging from mass services to highly professionalized and customized services. Second, the banking industry underwent several structural changes over the last two decades, supplementing branch-office services with a variety of digital services, which places high requirements on the employees. Third, due to the structural transformation mainly caused by digitalization, the workforce of this industry has been reduced considerably by cutting down branch networks. Thus, personal services and customer relationships become more and more important to remain competitive.

Since the case study is limited to one particular industry, the findings may not be fully generalizable to the whole service industry. Thus, it may be subject for future research to extend this study to other service industries in order to validate the results.

1.5 Structure of the Thesis

The thesis includes nine chapters. The figure below shows the structure of the thesis.

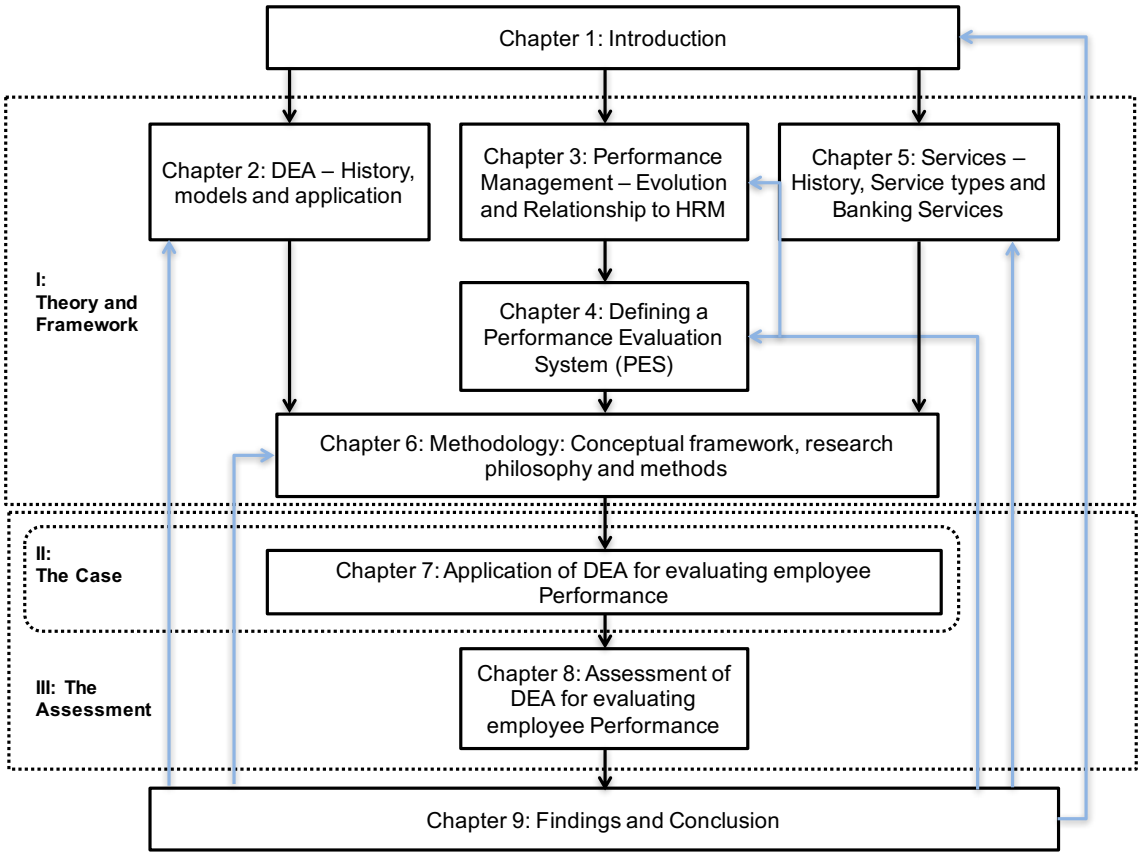


Figure 1: Structure of the thesis

Part one deals with the theoretical foundation reviewing research areas that are mainly related to the thesis´ topic, namely Data Envelopment Analysis, HRM and Performance Management, Service Management and the Research Philosophy and Methodology of this research. Thus, in chapter 2, the history and the theoretical foundations of DEA are illustrated and some important extensions to the basic model are introduced. Furthermore, the chapter provides an overview of the various fields of DEA applications during the last decades. The focus of chapter 3 is the evolution of Performance Management and its relation to Human Resources Management (HRM). Along with the definition of terms, Performance Management and its associated evaluation methods are discussed and the impact of performance evaluation on actual performance is investigated. Subsequently, the development of a comprehensive Performance Evaluation System, including the definition of elements and purposes, is outlined in chapter 4. Chapter 5 illustrates the evolution of Service Management and discusses the question whether there are unique characteristics that distinguish services from goods.

Furthermore, banking services are analysed with respect to their service characteristics. Eventually, challenges for managing banking services are derived. From the literature review, a conceptual framework is developed in chapter 6. Moreover, chapter 6 illustrates the research philosophy and the underlying methodology of this research.

Based on the theoretical foundation in part one, part two demonstrates DEA´s application in a German Cooperative Bank and reflects on the results of the subsequent assessment with all stakeholders. In this respect, chapter 7 illustrates the application of the DEA methodology, including the background and the purpose of the investigation as well as the results analysis. Based on this, recommendations for HRM will be discussed on individual and organizational level. Finally, in part three, the assessment of the proposed DEA-based approach by all relevant stakeholders is presented in chapter 8.

Concluding, chapter 9 sums up the major findings from the DEA application and the subsequent assessment. The findings are discussed and implications for both theory and practical application are provided. The findings shall provide several insights on the applicability of DEA on individual level, but also on the requirements for evaluating employee performance in the service sector and general issues with performance evaluation. Therefore, they may reference back on theory and framework. Furthermore, limitations and need for further research are pointed out.

2 Introducing Data Envelopment Analysis (DEA)

Some thirty years after DEA was introduced, it has become a well-known method in operations research (Cook and Seiford, 2009) and underwent an impressive development concerning both practical applications and theoretical advancement. In this chapter the history and the theoretical foundations of DEA are illustrated and some important extensions to the basic model are introduced. Furthermore, concepts for sensitivity analysis of DEA are discussed along with the advantages and disadvantages of DEA compared to alternative methods for measuring productivity. The final section of this chapter provides an overview of the various fields of DEA applications during the last decades.

2.1 The history and rationale of DEA

The Data Envelopment Analysis (DEA) is a non-parametric linear programming approach that evaluates the efficiency of a set of entities called “decision making units” (DMU). Due to its non-parametric nature, DEA requires no a priori specifications of the parametric form of the production correspondences (Banker and Morey, 1986a). It is able to handle multiple inputs and output figures simultaneously in complex operating situations and to provide a single efficiency measure (score) as a result.

Charnes, Cooper and Rhodes (1978) first introduced DEA - which is based on the earlier work of Farrell (1957) – in 1978. By proposing an activity analysis approach, Farrell´s intention was to develop a superior method for evaluating productivity. He drew his attention to the “Pareto-Koopmans optimality” noting that no final good is to be improved if this improvement results in worsening one or more final goods (Koopmans, 1951). Farrell (1957) extended this concept to inputs and outputs by eschewing the use of prices or other exchange mechanisms at the same time. He also utilized the performance of other DMUs to analyze each DMU relative to the input and outputs of other DMUs. The resulting measure is referred to as “Farrell measure of efficiency”. However, Farrell´s work was not able to account for multiple outputs and could not be applied to a large set of data. At this point Charnes, Cooper and Rhodes (1978) extended Farrell´s work by developing a dual pair of linear programming problems (see section 2.2), enabling the use of multiple inputs and outputs in ways that could locate the amount of inefficiencies for each DMU (Cooper et al., 2011). Their work originated in the non-profit sector by evaluating the efficiency of an experimental educational program called “Program Follow Through”. Due to the non-profit background the productive outputs were not of a type that could be reduced to a single “profit measure”. Although DEA has also been applied

to the profit-oriented enterprises hereafter, the more general term “decision making unit” rather than “business unit” is still commonly used.

The method itself is based on a linear programming problem (LP) that is solved for each DMU within a set of units. Each LP uses the input/output values of all other units under consideration. Thus, an efficient DMU can be identified by comparing it to each DMU within the set (Golany et al., 1990). This comparison can be carried out over all relevant inputs and outputs as DEA is invariant concerning units of measurement. It is able to accommodate measures as long as they are quantifiable. For this reason DEA *“may also assume a variety of forms which admit only ordinal measurement e.g. psychological tests, arithmetic scores, psychomotor skills”* (Charnes et al., 1978) and therefore is also able to account for inputs and outputs of qualitative nature. By transforming all inputs and outputs into a scalar measure, DEA computes a single score θ for each DMU, which in traditional DEA literature is referred to as “efficiency score”. However, contemporary authors prefer to use the term “performance score” (Cook et al., 2014), which will be applied throughout this thesis.

Due to its non-parametric nature, DEA computes the weights (multipliers) for each DMU on an individual basis. Although this liberalism does not necessarily lead to an efficiency of 100 per cent this principle allows each DMU to “shine in its best light” (Doyle and Green, 1994). A DMU is deemed to be efficient (performance score of $\theta = 1$) if its output is optimal for its inputs in comparison to inputs and outputs of all other DMUs under consideration. Thus, the efficient DMU cannot increase its output without increasing one or more inputs or decreasing another output, which again refers to the Pareto Optimality-requirement (Donthu, Yoo 1998). The efficient DMUs span a so-called efficient frontier. DMUs whose performance score θ is less than one are placed inside the frontier. Hence, they are “enveloped” by the frontier. For each DMU that is not on the frontier, DEA identifies a set of references that operate with similar input-output combinations, but manage to be efficient. Those references serve as peers for the inefficient DMU and can help to identify sources of inefficiency. Furthermore, they provide information on the extent of inefficiency.

Although DEA does not require a priori assumptions about the analytical form of the production function, some axiomatic formulations are required (Seiford and Thrall, 1990). Considering these axioms, the production function is obtained directly from observational data, which also has the advantage, that peers are existing DMUs and not a theoretical average.

To illustrate this, a simple model is constructed with input-output configurations observed for each of

j = *1,…. n* DMUs as (X_j, Y_j)

where X_j= $(x_{1j},\ldots\ldots x_{mj})$ is a vector of observed inputs of a DMU(j)

and Y_j= $(y_{1j},\ldots\ldots y_{sj})$ is a vector of observed outputs of a DMU(j)

It is assumed that each DMU(j) consumes the same inputs and produces the same outputs - in varying amounts. Additionally, it is assumed that at least one input and one output is positive. Subsequently the production possibility set T is determined by T = $\{(X,Y)|Y \geq 0$ can be produced from $X \geq 0\}$.

Given this production possibility set, the following properties are postulated (Banker et al., 1984)

1) No output without input: To produce a certain amount of output, a certain amount of input has to be consumed.
2) Convexity: If $(\mathrm{X_j}, \mathrm{Y_j}) \in \mathrm{T}, \mathrm{j} = 1, \ldots\ldots, \mathrm{n}$ and $\lambda_\mathrm{j} \geq 0$ are nonnegative scalars, then$\left(\Sigma_\mathrm{j}\lambda_\mathrm{j}\mathrm{X_j}, \Sigma_\mathrm{j}\lambda_\mathrm{j}\mathrm{Y_j}\right) \epsilon\, \mathrm{T}$. That implies that all (virtual) DMUs that result from input-output combinations of existing DMUs are part of the production possibility set
3) Free disposability (inefficiency postulate): If $(\mathrm{X}, \mathrm{Y}) \in \mathrm{T}$ and $\overline{\mathrm{X}} \geq \mathrm{X}$, then $(\overline{\mathrm{X}}, \mathrm{Y}) \in \mathrm{T}$.If $(\mathrm{X}, \mathrm{Y}) \in \mathrm{T}$ and $\overline{\mathrm{Y}} \geq \mathrm{Y}$, then $(\mathrm{X}, \overline{\mathrm{Y}}) \in \mathrm{T}$. Inefficient DMUs that consume a larger amount of input or produce a smaller amount of output are always part of the production possibility set.
4) Ray unboundedness: If $(\mathrm{X}, \mathrm{Y}) \in \mathrm{T}$ then $(\mathrm{kX}, \mathrm{kY}) \in \mathrm{T}$ for all $\mathrm{k} > 0$. Assuming a constant efficiency level, an increase/ decrease in input leads to a proportionate increase/ decrease in output.[1]
5) T is the intersection set for all $\widehat{\mathrm{T}}$ satisfying postulates 1 to 4, and subject to the condition that it includes all of the observed vectors $(\mathrm{X_j}, \mathrm{Y_j}) \in \mathrm{T}, \mathrm{j} = 1, \ldots\ldots, \mathrm{n}$

Those postulates lay down low requirements to the construction of the “efficient frontier” and therefore enable an adaption of the production function based on observed data in a very flexible way.

The next chapter introduces one of the most common DEA models: the CCR-model, which was proposed by **C**harnes, **C**ooper and **R**hodes in 1978. To illustrate the use of

[1] This does not account for the variable returns to scale model.

DEA an example of a single input and output case and examples of multiple inputs or outputs are provided. The measures are limited to multiples of two to enable a graphical illustration.

2.2 The basic CCR model

In general, the evaluation of performance can take a variety of forms, e.g. costs per unit, profit per unit or customer satisfaction per unit. Those evaluations are commonly stated in ratios like

$$\frac{Output}{Input}$$

This ratio is often used to measure productivity. In the case of one output and one input this seems a rather trivial matter. If multiple inputs and/ or outputs are used, a method for aggregating these factors must be applied in order to obtain a single ratio measure. The term “productivity” used in this context is referred to as “total factor productivity”, involving all factors of production. On the contrary, traditional measures of productivity, e.g. labour productivity in a company or land productivity in farming, are referred to as “partial productivity measures”. When considered in isolation, measures of partial productivity may indicate a misleading understanding of productivity (Coelli et al., 2005). The term “efficiency” is often used as a synonym for total factor productivity, but they are not precisely the same. The general understanding of “efficiency” is based on the concept of Pareto-Koopmans-optimality, which implies that efficiency is rather an evaluation of units relative to other units under consideration, than to absolute standards.

Based on those considerations, Koopmans (1951) defined the requirements to an efficient situation as follows:

” [A unit] is efficient whenever an increase in one of its coordinates can be achieved only at the cost of a decrease in some other coordinate”.

Thus, the relationship between productivity and efficiency is widely discussed in literature. Fried, Lovell et al. (2008) describe the relationship between the two terms as follows: productivity simply is the ratio of output to its input or, in terms of total factor productivity, the ratio of two scalars. Thus, productivity growth is the difference between output growth and input growth. Efficiency, on the other hand, is a comparison between observed and optimal values for inputs and outputs. In this context, the optimum is defined in terms of production possibilities (technical efficiency) or value terms (economic efficiency) (Fried et al., 2008). In consequence, a unit may be technical

efficient, but may still be able to increase its productivity by exploiting scale economies (Coelli et al., 2005). A similar definition is provided by Jacobs, Chase et al. (2004) who state productivity is the ratio of outputs over inputs, whilst efficiency is "the ratio of actual outputs to standards". This implies that efficiency is a relative concept.

Charnes et al. attempted to measure efficiency involving multiple inputs and outputs by encountering difficulties such as choosing suitable input and output measures or assigning weights to the chosen measures. By utilizing mathematical programming techniques, Charnes et al. (1978) designed a model (known as the CCR Model) that:

"...generalized the single-output/input ratio measure of efficiency for a single DMU in terms of a fractional linear-programming formulation transforming the multiple output/input characterization of each DMU to that of a single "virtual" output and virtual input." (Charnes, 1994)

They defined the efficiency of an arbitrary unit j as the weighted sum of j´s output over the weighted sum of j´s input. This is mathematically expressed by:

$$efficiency\ of\ DMU\ (j) = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \qquad 1.1$$

Consider a set of *n* DMUs, where each DMU(j) (j= 1, ….n) consumes *m* inputs *x(ij)* (i= 1,2,…m) and generates *s* outputs *y(rj)* (r= 1,2….s). The inputs are assigned to the weights (or multipliers) *v(i)* and the outputs to the weights *u(r)*. Additionally it is assumed that $x_{ij} \geq 0$ and $y_{ij} \geq 0$ and that each DMU has at least one positive input and one positive output. To evaluate the efficiency of a particular DMU (DMU(j) = DMU(0)) Charnes et al. (1978) introduced the "ratio form" of DEA. In this form the ratio of inputs to outputs is used to measure the relative efficiency of DMU(0) by evaluating it relative to the ratios of all other DMUs under consideration. The result obtained in this manner can be interpreted as the reduction of the multiple outputs and inputs to a single "virtual" output and "virtual" input.

For a certain DMU(0) the ratio of this single virtual output to the single virtual input provides a single measure of efficiency (performance score θ). The ratio for DMU(0), which is to be maximized, can be expressed by:

$$\max h_0(u, v) = \frac{\sum_{r=1} u_{r0} y_{r0}}{\sum_{i=1} V_{i0}\, x_{i0}} \qquad 1.2$$

As 1.2 is unbounded, a set of normalizing constraints (for each DMU) reflect the condition that the virtual input/output ratio for each DMU is less than or equal to unity (Cooper et al., 2011) and the multipliers are non-negative. This may be stated as:

$$\max h_0(u,v) = \frac{\sum_{r=1} u_{r0} y_{r0}}{\sum_{i=1} V_{i0}\, x_{i0}}$$

$$s.t.\ h_j = \frac{\sum_{r=1} u_{r0} y_{rj}}{\sum_{i=1} V_{i0}\, x_{ij}} \leq 1, \forall j = 1, \ldots\ldots, n \qquad 1.3$$

$$u_{r0} \geq 0;\ v_{i0} \geq 0; r = 1, \ldots, s\ ; i = 1, \ldots\ldots, m$$

There is no need for a-priori assumptions about the multipliers. All DMUs that achieve a performance score of one are deemed as relative efficient. Since the variables comprise only observed input and output values, DEA produces only relative efficiency measures. The performance score is calculated in relation to all other DMUs, using the actual observed values for inputs and outputs for each DMU (Charnes, 1994). A DMU is inefficient, if there is no set of multipliers with whom the DMU could achieve a performance score of one. Therefore, all inefficient DMUs obtain a performance score less than one (for the input-oriented case). In (1.3.) both the objective function and the constraint are a ratio of two linear aggregations. Since this ratio yields an infinite number of solutions (i.e. if (*u, v)* is optimal then *(αu, αv)* is also optimal for all *α>0)* the transformation developed by Charnes, Cooper and Rhodes for linear fractional programming selects a solution (i.e. the solution for (*u,v)* for which $\sum_i^m v_i\, x_{i0} = 1$). This yields the equivalent linear programming problem, which is referred to as CCR (Charnes, Cooper, Rhodes) model. Due to the transformation the multiplier variables are changed from *(u,v)* to (μ, v) (Cook et al., 2014). By applying the CCR transformation one obtains the following objective function:

$$\max z = \sum_{r=1}^{s} \mu_r\, y_{r0} \qquad 1.4$$

By the CCR transformation the objective function is being linearized which is due to several constraints. To maximize the numerator and to prevent unbound solutions, the denominator is set equal to a constant (say 1):

1.5

$$s.t. \sum_{i=1}^{m} \nu_i \, x_{i0} = 1$$

Further, to prevent DMUs from selecting weights that would cause the unit to obtain an performance score greater than one, the weighted sum of the units´ outputs must be less or equal to the units´ sum of weighted inputs with all multipliers being non-negative:

$$\sum_{r=1}^{s} \mu_r \, y_{rj} - \sum_{i=1}^{m} \nu_i \, x_{ij} \leq 0 \; \forall j = 1, \ldots . n \qquad 1.6$$

$$\mu_r, \nu_i \geq 0 \; for \; all \; r, i$$

Since the constraint requests that the aggregated output must not exceed one, the efficiency of DMU(0) will either equal one in which case it is efficient relative to the other units or will be less than one in which case the unit is inefficient. The weights, that are being calculated this way, can be interpreted as shadow prices. Thus, (μ, ν) are the set of most favourable weights for DMU(0) in order to maximize the numerator. In this sense, ν_i is the optimal weight for input item (i), while μ_r is the optimal weight for output item (r). In this manner the linear programming model determines a set of optimal weights for each DMU and therefore is referred to as "multiplier model". The performance score θ is determined by comparison *of* DMU(0) to all other (n) DMUs within the set. Hence, one version of the above model must be solved for each unit in turn.

By duality, this maximization problem can be transferred into a dual minimization problem (Cook et al., 2014):

$$\min \; \theta$$

$$s.t. \sum_{j} \lambda_j x_{ij} \leq \theta_0 x_{i0}, i = 1, \ldots . , m \qquad 1.7$$

$$\sum_{j} \lambda_j y_{rj} \geq y_{r0}, r = 1, \ldots . , s$$

$$\lambda_j \geq 0 \; j = 1,2, \ldots \ldots n$$

Variable θ is identical to variable z . θ indicates the DMU´s distance to the efficient frontier. The multipliers λ_j, that stem from the transfer into a dual minimization problem,

show which other DMU may serve as references for DMU(0) and provide information about their emphasize (if DMU(0) is efficient then DMU λ_0 = 1 and all λ_J = 0) (Westermann, 1996) .Model, (1.7.) is referred to as “envelopment” model. Thus, it spans the efficient frontier. By virtue of the dual theorem of linear programming, either model can be used to obtain an performance score (Cooper et al., 2011).

If one addresses the output side and considered the ratio of virtual input to virtual output, the objective has to be reoriented from maximization to minimization to obtain:

$$\min h_0(u,v) = \frac{\sum_{i=1} v_{i0} x_{i0}}{\sum_{r=1} u_{r0}\, y_{r0}}$$

$$s.t.\ h_j = \frac{\sum_{i=1} v_{i0} x_{i0}}{\sum_{r=1} u_{r0}\, y_{r0}} \geq 1, \forall j = 1, \dots\dots, n \qquad 1.8$$

$$u_{r0} \geq 0;\ v_{i0} \geq 0; r = 1, \dots, s\ ; i = 1, \dots\dots, m$$

Applying the “Charnes-Cooper” transformation to this problem, one obtains the multiplier model (1.8) and – by duality – the envelopment model (1.9):

$$\min q = \sum_{i}^{m} v_{i0}\, x_{i0}$$

$$s.t. \sum_{r}^{s} \mu_r\, y_{r0} = 1 \qquad 1.9$$

$$\sum_{i}^{m} v_i\, x_{ij} - \sum_{r}^{s} \mu_r\, y_{rj} \geq 0\ \forall j = 1, \dots\dots n$$

$$\mu_r, v_i \geq \varepsilon\ for\ all\ r, i$$

$$\max \varphi$$

$$s.t. \sum_{j} \lambda_j y_{rj} \leq \varphi_0 y_{r0}, r = 1, \dots\dots, s \qquad 1.10$$

$$\sum_{j} \lambda_j x_{ij} \geq x_{i0}, i = 1, \dots\dots, m$$

$$\lambda_j \geq 0\ j = 1,2, \dots\dots n$$

Comprehensive examples for the use of DEA for single and multiple input and output case are provided by Cooper et al. (2007).

2.3 Extensions to the basic model

The previous chapters illustrated the development of the basic DEA model – known as the CCR model. However, since first introduced in 1978, DEA has been refined and enhanced by a number of models. Five of these extensions will be discussed in this section. The models have been selected out of a great variety of extensions because of their significance in DEA literature on the one hand. On the other hand, they have been chosen due to their relevance for applying DEA to measure employee performance on an individual level.

2.3.1 Slacks and the non-Archimedean element

The diagram below points out a major problem concerning Pareto-optimal performance:

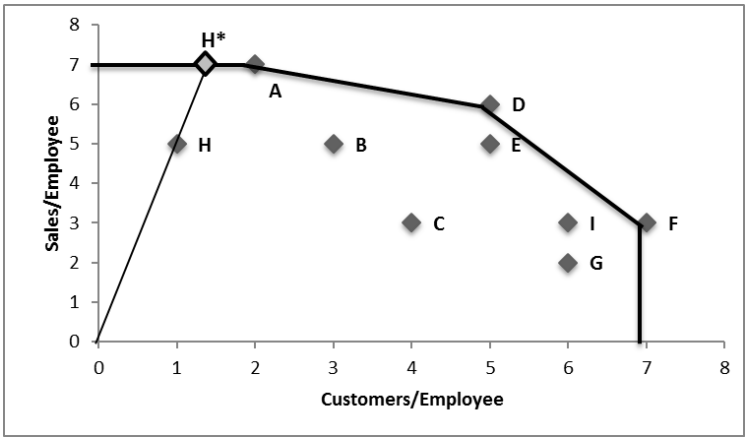


Figure 2: slacks in DEA

To become "radial efficient" DMU H has to increase its outputs proportionally to attain point H* on the efficient frontier. However, in the sense of Pareto-optimality, there is still room for improvement when reaching point H* by increasing the output "customers/employee" without worsening the other output. This is called "output-slack", which is expressed by the distance $\overline{H^*A}$. Only after the output "customers/employee" has separately been increased to reach point A, DMU H can be considered Pareto-efficient. Thus, radial-efficiency is no indicator for Pareto-efficiency and units such as point H* are referred to as "weakly efficient". To become Pareto-efficient the value of non-radial improvement of the output "customers/employee" has to be added to the value of radial

improvement. Based on those considerations, Charnes, Cooper and Rhodes (1978) enhanced the basic model by introducing so called slack variables (input-slacks = s^-, output-slacks = s^+). If there are positive slacks, the slack values have to be added to the radial values for improvement (Bauer et al., 2015). Additionally, a "non-Archimedean" element ε is added, which is defined to be smaller than any positive real number (without specifying the value). This element prevents that parts of the efficient frontier have a slope of 0 or infinity.

The input-oriented CCR model is developed as follows:

$$\begin{gathered} \min\ \theta - \varepsilon \left(\sum_r s_r^+ + s_r^- \right) \\ s.t. \sum_j \lambda_j x_{ij} + s_i^- = \theta_0 x_{i0}, i = 1, \ldots\ldots, m \\ \sum_j \lambda_j y_{rj} + s_i^+ = y_{r0}, r = 1, \ldots\ldots, s \\ \lambda_j, s_i^-, s_r^+ \geq 0\ \forall i, j, r \end{gathered} \qquad 1.13$$

By this, a DMU is Pareto-efficient (or "fully efficient") if $\theta = 1 \wedge s^+ = 0 \wedge s^- = 0$. A DMU with $\theta = 1$, but with at least one slack being positive therefore is weakly efficient.

In an attempt to address slacks, Charnes et al. (1985) developed the "Additive model" that directly deals with input surplus and output shortages. Although the model can discriminate between efficient and inefficient units by the existence of slacks, it is not able to gauge the depth of inefficiency. In order to calculate a scalar measure, several authors proposed models to define inefficiency based on slacks. Among them are Lovell and Pastor (1995), Thrall (1996) and Tone (2001).

2.3.2 Variable returns to scale

In literature of traditional economics, returns to scale (RTS) have mostly been analyzed for single-outputs situations, with increasing RTS being defined as the increase in input resulting in a more than proportional increase of output (Banker et al., 2011). The traditional CCR model does not account for RTS, although in some situations it seems unsatisfactory to ignore the influence of RTS. Thus, Banker et al. (1984) and Banker and Thrall (1992) developed a DEA model to incorporate RTS for multiple outputs, assuming that there are best-practice cases for different scale levels (input volumes).

The diagram below illustrates a case with one input and one output. The straight line running through the origin represents the efficient frontier determined by the CCR model. The only efficient DMU in this case is DMU A.

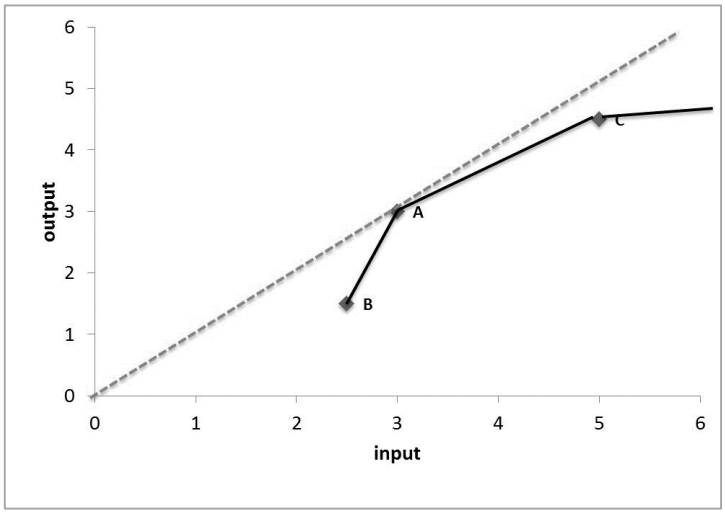


Figure 3: variable returns to scale

If there would be a non-linear relationship (e.g. in the case of variable returns to scale), DMUs B and C could additionally establish the frontier line. Thus, the frontier that is formed by B, A and C considers inefficiencies that are caused by scale size. DMU B is located in the area of rising returns to scale. Therefore, an increase in input would lead to a disproportionate increase in output. DMU C, on the other hand, is located in the area of decreasing returns to scale with an increase in input leading to an under-proportionate increase in output. Hence, B and C also produce their outputs efficiently but their size is either be too small (B) or too large (C). Whether a DMU possesses the optimal size is indicated by the λ_j-values that are calculated by the CCR-model (envelopment model, see section 2.2):

$\sum_j^J \lambda_j = 1$ DMU has optimal size

$\sum_j^J \lambda_j > 1$ DMU is too large

$\sum_j^J \lambda_j < 1$ DMU is too small

Thus, the efficiency of a DMU assuming variable returns to scale can be evaluated by applying the BBC model, named after Banker, Charnes and Cooper:

$$\min\ \theta - \varepsilon\ (\sum_r s_r^+ + s_r^-)$$

$$s.t. \sum_j \lambda_j x_{ij} + s_i^- = \theta_0 x_{i0}, i = 1, \ldots\ldots, m \qquad 1.14$$

$$\sum_j \lambda_j y_{rj} + s_i^+ = y_{r0}, r = 1, \ldots\ldots, s$$

$$\sum_j^J \lambda_j = 1$$

$$\lambda_j, s_i^-, s_r^+ \geq 0\ \forall i, j, r$$

The BBC-model enhances the CCR-model by the constraint $\sum_j^J \lambda_j = 1$ ensuring that only DMUs with a similar level of scale are compared to one another and economies of scale are being removed from the calculation of the performance score (Bauer et al., 2015).

In literature there are also two major points of criticism concerning the application of the BBC model. One point of criticism is, that it is not certain whether inefficiencies calculated by the CCR model are really results of RTS effects. Another point is, that the application of the BBC model increases the number of efficient DMUs, which affects the significance of a DEA analyses. Therefore, empirical studies often fall back to the CCR model (Westermann, 1996).

2.3.3 Incorporating uncontrollable variables

Given the rather complex nature of services, DEA seems to be a suitable instrument for assessing employee performance and setting targets, as it is able to accommodate both controllable and uncontrollable factors (Moreno and Tadepalli, 2002). Uncontrollable factors are usually environmental or competitive factors that are outside the DMU´s control but nonetheless affect their performance, e.g. location or clientele in the area (Donthu and Yoo, 1998). Therefore, many authors use the term “environmental variable” synonymously. DEA literature classifies two main approaches to account for uncontrollable variables:

a) directly incorporating uncontrollable factors into DEA models
b) using multi-stage approaches

The first mentioned approach is based on the works of Banker and Morey (1986a; 1986b), who introduced two DEA models that deal with uncontrollable factors, depending

on whether they are of non-discretionary or categorical nature. An illustrative example for an uncontrollable non-discretionary variable in the banking sector is the population within the catchment area (exhibiting characteristics such as income, age or other demographic factors). This variable is a key factor in order to obtain a given level of deposits or loans, but it lies outside the control of the employee. A major drawback of the non-discretionary model is the prerequisite to know the direction in which the uncontrollable variable is going to influence the efficiency. Additionally, as the model is not able to remove all environmental effects, it tends to overestimate the efficiency of DMUs with more desirable conditions (Paradi et al., 2010).

An enhancement to the non-discretionary model is the categorical model. In this case the uncontrollable variable is of categorical nature (e.g. the existence or non-existence of a car park or cash machines). To enable the incorporation of categorical variables, Ruggiero later extended the Banker and Morey´s model ((1998). In the case of categorical variables, the DMUs are grouped into categories, with the model assuming that there is a hierarchy in categories. To avoid that the category is treated as it had cardinal meaning, (e.g. a DMU in category 4 is not automatically twice as important as a DMU in category 2) the reference set is restricted by excluding the units with more desirable environments. Thus, only DMUs that are in the same or lower categories are compared to each other. A prerequisite of the categorical approach subsequently is, that the size of a category is large enough to maintain a satisfactory level of discrimination between the DMUs (Paradi et al., 2010). The mathematical derivation of both models can be found in Banker and Morey (1986b).

Another main approach to incorporate uncontrollable variables in DEA is a multi-stage method, first proposed by Timmer (1971) using a two-stage approach. In the first stage of the approach a basic DEA model is applied to assess the efficiency of all DMUs in the set, without considering uncontrollable variables. In a second step, the performance scores are regressed against each uncontrollable variable to evaluate and separate its impact. Since first proposed, many extensions have been made to this approach (e.g. enhancing it to a three- or four-stage approach). Nethertheless, a crucial drawback to the multi-stage approach is that the second stage requires an a priori specification of the functional form – and with this contradicts a main advantage of DEA (Paradi et al., 2010). Although both approaches have their weak points, the Banker and Morey models are less complex and easier to handle. Hence, to develop a method for performance measurement that is suitable for practical application in companies, those models seem more appropriate. Additionally, their flaws can be addressed by performing a sensitivity analysis.

2.3.4 Incorporating qualitative data

Although in Performance Measurement Research there is still no consent about the question whether it is more suitable to use input or output measures for performance evaluation (C. E. Pettijohn et al., 2001), it is widely agreed that the utilized method should account for qualitative data as well as for quantitative data (K. Becker et al., 2011; Sudin, 2011). Jackson et al. (2010) provided a comprehensive overview stating that particularly in performance evaluation of sales forces qualitative data have mostly been used as input data. This may be explained through the fact that input data often comprise behavioral factors, such as communication skills, initiative or attitude (L. S. Pettijohn et al., 2001). Other studies show that personal factors like education, cultural skills or ratings in an interview can be included as an input as well (A. C. Soteriou and Stavrinides, 1997; Warning, 2014). Warning (2014) used a DEA approach to include job interview ratings as an input variable for staff selection. On the other hand, there are many DEA studies that use qualitative data as an output, especially to address customer orientation or satisfaction. Soteriou and Stavrinides (1997) applied DEA for the purpose of developing a service quality framework using the obtained level of service quality as an output. Wagner et al. (2003) introduced the use of DEA for physician profiling using patient surveys of a data source for customer satisfaction. Bayraktar et al. (2012) applied DEA to measure the efficiency of customer satisfaction and loyalty for mobile phone brands.

An issue that always arises in dealing with qualitative data - no matter if it is used as an input or an output measure - is that this data is often difficult to quantify. Thus, qualitative factors are often captured either on Likert-Scales or are represented in rank positions in an ordinal rather than in a numerical sense (Cook et al., 1996). As mentioned before, the traditional DEA model treats rank order information as if it had cardinal meaning. For this reason, Cook et al. (1996) developed a model to enable the use of ordinal and cardinal factors in DEA. In 1999, Cooper et al. provided a model for the same purpose in the form of the imprecise DEA: IDEA (Cooper et al., 1999). Later, Cook and Zhu (2006) evaluated the existing models and proved their equivalence. They showed that both approaches can be transformed into a traditional VRS (CRS) model.[2] Therefore, it is possible to include qualitative data in conventional DEA models.

[2] The discussion of the approaches and the mathematical transformation of the models can be found in Cook (2011).

2.3.5 Restriction of weights

One of the major advantages of DEA models is, that there is no need for a priori specification of input and output weights. Efficient DMUs are identified by letting the weights "run freely", which allows each DMU to choose its own weights and to show itself in its best light (Schaffnit et al., 1997). Although this characteristic is an attractive one, it may lead to extremes. Thus, a DMU might achieve 100 per cent efficiency by weighting only a single input and a single output assigning the rest with zero weights or as Doyle and Green (1994) put it: *"Best engine on the market – pity that the car has no wheels."* Thanassoulis (1995) argued in a similar way in his study to assess police forces. He felt that clearing up a violent crime is more valuable than clearing up a burglary and therefore weighted the first case ten times higher than the latter. There are also other reasons to add weight constraints. Paradi and Schaffnit (2006; 2004) incorporated managerial information for they felt that unconstrained DEA results are contrary to managers´ views and beliefs or were quite unrealistic. Additionally, managers´ priorities (e.g. cost reduction) could be reflected more properly. By prohibiting extreme weighting divergences the total number of efficient units usually drops (Cooper et al., 2011; Thanassoulis, 1995). Thomas et al. (1998) made use of this effect. In order to assess retail store efficiency, they assigned weights to inputs and outputs. Thus they were able to consider a larger number of input criteria.[3] Pedraja-Chaparro et al. (1997) argue, that if the inputs or outputs, that are given zero weight in the analysis, were important, the measure of efficiency would be deficient in placing no consequence on them. If they were not important, on the contrary, then why were they included in the analysis in the first place?

There are two main approaches for incorporating weighting constraints in DEA. For the purpose of identifying the best site for the location of a high-energy laboratory, Thompson et al. (1986) developed the "assurance region approach". This approach imposes constraints on the relative magnitude of the weights for special items. To evaluate bank performances Charnes and Cooper (1990) provided another approach by developing the so called "cone-ratio envelopment". The theoretical aspects of the approaches as well as their various applications can be found in Allen et al.(1997) and Thanassoulis et al.(2004). However, the discussion to impose weight restrictions is not without controversy in DEA literature. Some authors defend a straight "hands-off" policy, arguing that this is one of the major advantages of DEA. Therefore, the incorporation of weights should be well-argued.

[3] As rule of thumb, the total number of input and output measures should be less than one third of the total number of DMUs within the set.

2.4 DEA Sensitivity analysis - the problem of outliers

As the discussion about imposing weights externally has shown, DEA tends to be sensitive to outliers that achieve efficiency by rare input-output combinations. However, the freedom to choose its own weights is one of the strongest selling points to DMUs that are being appraised by DEA (Doyle and Green, 1994). The traditional DEA groups DMUs into two general sets: Efficient DMUs that define the efficient frontier and inefficient DMUs. Efficient DMUs all receive a first place in the ranking. However, beyond this dichotomized classification, decision makers are often interested in a complete ranking in order to refine the assessment. Thus, a drawback of DEA that is often discussed in literature is its lack in discrimination of efficient units in cases when the number of inputs and outputs is too high relative to the number of units (N. Adler et al., 2002). To overcome problems that occur in the practical use of DEA, several concepts have been developed to enable the identification of outliers without being forced to add additional constraints and to rank efficient DMUs. In the following the concept of superefficiency, which is a procedure to rank efficient DMUs by comparing them to a reference frontier spanned by all other DMUs. It was proposed by Andersen and Petersen (1993). The basic idea of the concept is to rerun the DEA model and removing, in turn, each efficient DMU (Wagner et al., 2003). The procedure enables an extreme efficient DMU *k* to achieve a performance score greater than one by using the primal model and removing the *k*th constraint:

$$\max \theta_k = \sum_r^s \mu_r \, y_{rk}$$

$$s.t. \sum_i^m \nu_i \, x_{ik} = 1 \qquad 1.17$$

$$\sum_i^m \nu_i \, x_{ij} - \sum_r^s u_r \, y_{rj} \geq 0 \; \forall j = 1, \ldots . . n, j \neq k$$

$$u_r, \nu_i \geq \varepsilon \; for \; all \; r, i$$

The score that is calculated for the efficient DMUs in this manner reflects the radial distance from the DMU under evaluation to the efficient frontier, estimated with the DMU excluded from the sample. The figure below illustrates the concept of superefficiency:

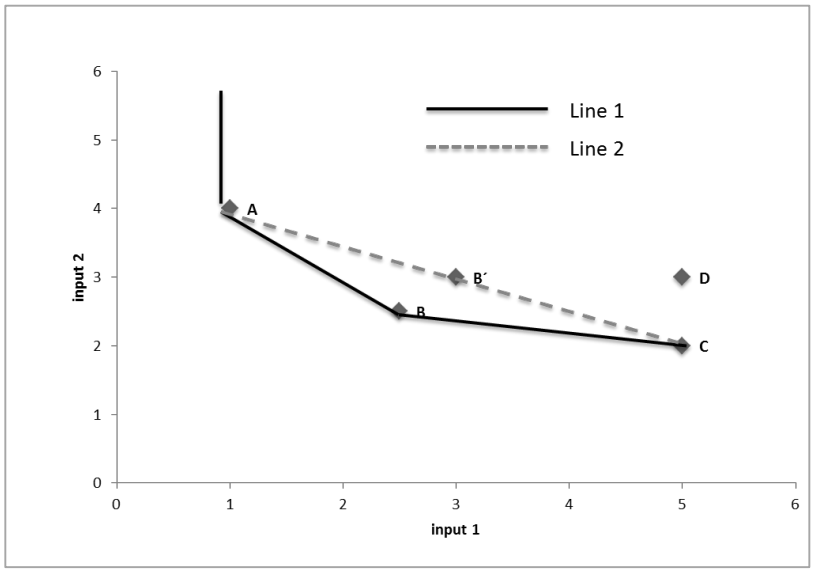


Figure 4: the concept of superefficiency

Consider an evaluation with two inputs (and the same output) with DMU A, B and C spanning the efficient frontier ("line 1"). If DMU B is removed from the calculation, the efficient frontier – as seen from DMU B´s point of view – moves away from the origin. Now, DMU B is projected to point B´ (6.0, 6.0) on the shifted efficient frontier ("line 2") constructed from all other observations in the sample excluding point B. Thus, the reference point for B becomes B´. As B has been assigned an performance score of 1, B´ is assigned a score of 1.2, which means that B may increase its inputs proportionally up to a factor of 1.2 and remain efficient (Andersen and Petersen, 1993). The score of superefficiency provides a measure of how much the efficient frontier shifts by the removal of an efficient unit and can therefore be interpreted as a measure of a DMUs influence (Wagner et al., 2003). Note that the exclusion of inefficient DMUs such as D does not affect the efficient frontier. Therefore, the concept of superefficiency does not change the performance scores of inefficient DMUs.

However, Dula and Hickmann (1997) and Seiford and Zhu (1999a) demonstrated that various DEA superefficiency models are infeasible. Thus, superefficiency evaluation may not always provide a complete ranking of efficient units. Another approach to identify outliers is the concept of cross-efficiency evaluation, which was first proposed by Sexton et al. (1986) and was re-examined by Dolye and Green (1994) in detail (Liu et al., 2016).

2.5 DEA Applications in the banking industry and on employee level

Since first proposed in 1978, DEA has been the focus of considerable research and has widely been applied to various fields of research and for various purposes. In their comprehensive analysis of 40 years of DEA literature, Emrouznajed and Yang (2017) found that from the first work of Charnes et al. in 1978 until 2007 more than 10,300 scholarly articles on DEA have been published in journals or as book chapters. Due to an easier access to corporate data, the majority of DEA applications are focused on institutional level (Paradi et al., 2011). DEA has been used to evaluate performances of a wide spectrum of organizations like hospitals (Du et al., 2014; Wagner et al., 2003) universities (B. L. Lee and Worthington, 2016; Nazarko and Šaparauskas, 2014; Tzeremes and Halkos, 2010), airlines (Fethi et al., 2000; Wanke and Barros, 2016) or even rice farms (Dhungana et al., 2004). Considering the papers published on DEA applications, it becomes evident that DEA has widely been applied in banking industry. According to Emrouznajed and Yang (2017) the banking industry was second most common application field in the years 2015 and 2016. However, within the banking industry it has not been applied on individual level, yet. Also in other fields there is very little experience in applying DEA on individual level.

In the following, an overview of DEA applications in the banking sector both on institutional and branch level will be provided. Subsequently, DEA studies that focus on performance measurement on employee level will be reviewed.

2.5.1 DEA applications in the banking industry on institutional and branch level

Although the banking sector is known to have dense branch networks most of the DEA studies in banking industry focus on institutional level. Berger and Humphrey (1997), who were the first to review efficiency analysis techniques applied to the banking sector, found that out of a total of 57 reviewed DEA applications only 15 where on branch level. In their comprehensive survey investigating the application of DEA in the banking industry, Paradi and Zhu (2013) found that between 1985 and 2011 a total of 275 studies on DEA applications in the banking industry have been published – only 80 of them on branch level.

Furthermore, the purposes of studies differ, depending on the decision whether to apply DEA to a bank as entity or to a branch as a unit. Based on the surveys of Berger and Humphrey (1997), Paradi and Zhu (2013) and Liu et al. (2016), the main application issues on institutional level are:

- To assess efficiency changes over time by assessing the effects of deregulation, mergers, or market structure on efficiency, e.g. Charnes et al. (1990) who apply DEA to U.S. commercial banks over a time period of 5 years or Barr et al. (1994) who examine bank failures over a period of four years.
- To compare the profitability and marketability, e.g. Seiford and Zhu (1999b) who used a DEA approach to compare the profitability of 55 U.S. commercial banks. Eken et al. (2014) applied a DEA-based risk and profitability approach to analyse the profitability of European banks.
- To improve bank performance by identifying “best practices” and “worst practices” and explaining efficiency differences, e.g. Sherman and Ladino (1995) who applied DEA to Mexican banks or LaPlante and Paradi (2015) using DEA to identify best and worst cases as benchmarks for the branch network of one of Canada´s top five banks.
- To accommodate international comparison, e.g. Thoraneenitiyan and Avkiran (2009), who investigate the relationship between post-crisis bank restructuring, country-specific conditions and bank efficiency in Asian countries from 1997 to 2001.

While DEA applications on institutional level cover a wide range of issues, the application on branch level mainly focuses on evaluating the branch´s specific operating aspects. The aim of those applications is mostly to identify and eliminate those deficiencies that are within the control of the branch´s management, such as the deployment of staff, cost efficiency or investigating the determinants of efficiency. The main research purposes of studies that applied DEA on bank branch level can be classified as follows:

- Profitability and cost efficiency analysis
- Efficiency rankings
- Branch intermediation analysis
- Environments and technology impacts on branch performance
- Effects of mergers and acquisition on branch performance

A comprehensive overview about studies dealing with the above mentioned research issues can be found in Paradi and Zhu (2013). Considering the purpose of measuring the performance of employees in the banking sector, there are studies on bank branch level that set some important preliminary foundations. Thus, Thanassoulis and Dyson (1992) developed a DEA model that enables the estimation of target levels in order to render inefficient units efficient. This was one of the first approaches that used DEA for target setting. Some years later Lovell and Pastor (1997) applied DEA to a Spanish bank

branch network for the purpose of setting performance targets and evaluating the procedure of target setting itself. Those studies proved that DEA has it´s practical use not only in identifying best practice, but also in constructing achievable and efficient performance targets.

Regarding the importance of service quality and customer satisfaction, there are also studies on bank branch level that developed strategies to incorporate those factors into performance analysis. Golany and Storbeck (1999) incorporated customer satisfaction as an output of financial services into their DEA model. A similar approach was used by Bayraktar et. al (2012) in order to investigate customer satisfaction and loyalty. The importance of perceived service quality for bank branch performance was also stressed by Soteriou and Zenios (1999), who used the SERVQUAL approach by Parasuraman et al. (1988) to measure customer satisfaction and incorporated that measure into a DEA model.

2.5.2 DEA Applications on individual level

In spite of its intense application on institutional and branch level and despite the importance of employee performance to the delivery of high quality services, there is still very little experience regarding the application of DEA on individual level. In addition, there is a lack of feedback information from managers´ and employees considering adequateness or comprehensibility of results. A first attempt to apply DEA on individual level aimed on sports teams. In 1992, Leibenstein and Maital introduced a DEA based approach to measure the x-inefficiency of an icehockey team with 19 members (Leibenstein and Maital, 1992), whereas Howard and Miller assessed the efficiency of baseball players (Howard and Miller, 1993). One of the first attempts to propose DEA on an individual level in a traditional company environment were Boles et al (1995), who applied the tool to measure the performance of salespeople of a firm selling advertises. Pilling and Donthu (1999) advanced that approach investigating the impact of territorial characteristics. In order to assess team performances Thanassoulis (1995) applied DEA to police forces in England. Paradi et al. (2006) used an advanced approach for engineering design teams of a Telecommunication firm in Canada. An approach to measure employee performance in the manufacturing industry was proposed by Manoharan et al. (2009). Another study against the background of the manufacturing industry was presented by Shirouyehzad et al. (2012), with the purpose of analysing motivation and job satisfaction rather than accounting for job performance. Warning (2014) developed a DEA based approach for personnel selection in a news agency.

The table below gives a comprehensive overview about DEA approaches that have been carried out on individual or group level:

	DMUs under assessment	**Input measures**	**Output measures**	**Remark**
Leibenstein and Maital (1992)	Hockey players	salary, shots on goal	goals per game, assists per game	1,2,4,5
Howard and Miller (1993)	Baseball players	29 variables including stolen bases, games played, run scores	salary	1,2,4,6
Boles, Donthu and Lothia (1995)	Sales forces of advertising firms	sales training, salary, management ration, territory potential	supervisor evaluation, sales volume	1,2
Thanassoulis (1995)	Police forces	Violent crimes, burglaries, other crimes, officers	Violent crimes clear ups, burglary clear ups, other crimes clear ups	1,2,3,4
Lovell and Pastor (1997)	Bank branch networks	17 target values including demand, deposits, home purchase deposits	per cent of target value actually achieved	1,2,3,7
Thomas et al. (1998)	Store managers	ratio of full-to-part-time employees, employee tenure, total inventory, average transaction size, employee turnover, inventory shrinkage	sales, profits	1,2

Pilling, Donthu and Henson (1999)	Sales representatives (Women´s and men´s apparel)	total market demand, avarage sales per account, growth rate of market demand	sales volume	1,2,6
Wagner, Shimshak and Novak (2003)	Physicians	12 input measures including number of member´s inpatient admissions, length of stay for inpatient admissions	panel, avarage weight, % of inpatients not readmitted to hospital within 15 days, % of inpatients with no complications, health status, quality score	1,2
Paradi, Smith and Schaffnit (2006)	Engineering design teams	EQE ("equivalent number of employees") Managers, EQE Associates, EQE Coordinators, total Dollar expenditures, held orders	total NAS ("network assessing services") gain, total line movements	1,2,3
Manoharan et al. (2009)	Employees in manufacturing industry (automotive industry)	years of experience, percentage of attendance, history of employee (interpersonal relations)	percentage of acceptable units produced, production volume	1,2,4,5

Shirouyehzad and Housseinzadeh (2012)	Pipe company employees	salary, physical working conditions, irresponsibility	motivation, job satisfaction, organisational commitment	1,2,4,8
Warning (2014)	Job applicants in a News Agency	Education, experience, language test score	Number of projects, volume of projects, job interview score	1,2,8

Table 1: DEA approaches on individual or group level

1: no consideration how managers´ evaluate results

2: no consideration how employees evaluate results

3: assessment on group-level or branch level

4: does not refer to service sector

5: very small number of units (<20)

6: just one output

7: applied for the purpose of target setting

8: does not assess performance in particular

Although some of the approaches were applied to employees working in the service industry, no study does explicitly account for the special requirements to measure performance in the service sector. Furthermore, there is a lack of information whether managers´ experienced the obtained results as helpful to base administrative decisions on (for instance promoting employees) or for executing developmental actions. Additionally, no study provides empirical data about employees´ responses towards the procedure and the results.

3 Introducing Performance Management as a domain of Human Resources Management

Performance Management is an essential part of the organization´s Human Resources Management (HRM) [4] activities. Although it is a widely applied and accepted management practice, there seems to be little consensus about its definition and the use of terms concerning its application on different organizational levels. Furthermore, due to radical changes in the workplace of the twenty-first century, the legitimacy of traditional HRM and HR practices has been questioned by many scholars.

Performance Management and its associated HR practices are supposed to have an impact on organizational and individual performance. In literature, however, the nature of the linkage between HRM, HR practices and organizational performance is a controversial issue. Along with the definition of terms and a critical evaluation of HRM, this issue will be discussed in the following chapter.

3.1 The practice of HRM

While this paper does not aim at investigating the complex topic of HRM, it is important to consider its core concept and terms in order to observe Performance Management as an essential part of HRM. Thus, the following section deals with the definition of HRM, the involved HR strategies and practices as well as recent developments concerning the strategical focus of HRM in the twenty-first century. Also, this sections addresses upcoming concerns and criticism on HRM and its related practices.

3.1.1 The essence of HRM

HRM is an area of varying definitions and contentious theory and its role within the organization is widely discussed in literature. Thus, the struggle of HR practitioners to establish their role has a long history. The development of the HRM concept has roots in the 1940s where it replaced the human relations approach to managing people founded by Mayo (1933). The supporters of the human relations approach believed that productivity was directly related to job satisfaction. As a concept, HRM was first

[4] In response to criticism of referring to people as "resources", the term "People Management" is sometimes preferred (Armstrong and Taylor, 2014). Due to its common use, the term "Human Resources Management" will be applied throughout this thesis.

mentioned in the 1960s but did not emerge in a comprehensive form until the 1980s were it was described in several frameworks like the "Harvard framework" or the "Matching model" mostly by US scholars and underpinned by a number of theories drawn from several academic fields like behavioural sciences but also strategic management (Armstrong and Taylor, 2014). Hence, HRM can be interpreted as a philosophy for managing people containing a number of guiding principles and beliefs on how this should be done. Particularly in the UK, the HRM concept has been heavily criticized during this time as being too manipulative and managerialist. Storey (2007), for example, argued HRM is "the potential manipulative nature of seeking to shape human behaviour."[5] In the 1990s, critical views particularly focused on "soft HRM" practices linking them to shaping employee subjectivity and "the production of seduced selves". The emphasis of the debate was put on new disciplinary strands, which – as postulated by Grant et al. (1998) – were "founded on the internalization of self-regulation, calculation and control in which externally imposed authority and discipline becomes less significant". As a result, HRM folded into a variety of disciplines and practices (P. Thompson, 2011).

Hence, the practice of HRM today is rather diverse and does not necessarily apply to the original concept. From a pragmatic point of view, HRM has rather become "something that organizations do instead of an aspiration or philosophy" (Armstrong and Taylor, 2014). Boxall et al. (2007) attest a "profound diversity" of HRM that covers "a vast array of activities (…) across businesses, units industries and societies". In terms of definition they describe HRM as "the management of work and people towards desired ends" and "a fundamental activity in any organization in which employees are employed". In a more recent study, Boxall (2013) argues that "human resources" should not be defined as the people that are employed in an organization. Rather, human resources should be understood as "the resources that are intrinsic to human beings" including knowledge, skills and energy as well as the underpinning dynamic characteristics of people like physical and emotional health, personalities and intellectual capabilities, which can be applied to various tasks in life. In consequence, individuals develop special human competencies, which are hard to copy and therefore may be a valuable source of competitive advantage for organizations. Thus, Boxall (2013) identifies the employment relationship as the "primary vehicle" for combining the needs of individuals and organizations. In essence, HRM is a comprehensive and coherent approach dealing with

[5] A detailed review of reservations about the original concept of HRM is provided by Armstrong and Taylor (2014).

the employment and development of people and their individual resources to support the organization achieving its objectives (Armstrong and Taylor, 2014) .

3.1.2 HR strategies and practices

To link people management policies and practices to the achievement of organizational outcomes, strategic HRM (SHRM) has emerged in the late 1970s an overarching concept and has widely been adopted since (Kramar, 2014). The underlying premise of SHRM is that HRM has to provide a strategic framework to support long-term business objectives. One of SHRM´s underpinning concepts is the resource-based view of the organization. According to the resource-based view, organizations can only achieve competitive advantage by creating value in a way that is rare and not easy to imitate (B. Becker and Gerhart, 1996; Messersmith et al., 2011). Whilst the role and design of HRM is still subject to scientific discourse, the idea that employees represent a key resource for organizational success is widely accepted. Especially in the service industry employees are of crucial importance due to the customer/ employee interaction, in which the customer engages in a subtle and complex experience of personal nature (Boudreau and Ramstad, 1998; Haynes and Fryer, 2000). Further, employees often are a source of unique skills and knowledge and therefore are a key driver for competitive advantage. Often it is easier for an organization to copy one another´s technology than to copy their human resource capabilities (Wall and Wood, 2005). For this reason, the resource "employee" not only is one of the most cited as a potential lever of sustainable competitive advantage, but is also believed to become the primary source of it within the next years (Messersmith et al., 2011; Wirtz and Ehret, 2017).

To generate a sustained competitive advantage and to improve organizational performance, SHRM adopts a bundle of HR practices which are interrelated and therefore reinforce each other (Takeuchi et al., 2009). In this respect, a general HR strategy describes the overall system or bundle of HR practices the organization puts into effect (Armstrong, 2010). To promote an HR strategy, the organization needs to identify and select appropriate HR practices. In this respect, the concept of a "High Performance Work System" (HPWS) including a bundle of "High Performance Work Practices" (HPWP) has emerged as a core construct aiming at the rigorous selection, management and retention of the best possible human capital (Takeuchi et al., 2009). Whilst there is no agreed list of HR practices that constitute an HPWS, incentive pay systems and performance management procedures are frequently cited practices (see (Albrecht et al., 2015; Delery and Doty, 1996; Messersmith et al., 2011; Purcell et al., 2003; Wall and Wood, 2005). By introducing HPWS practices organizations intend to

improve organizational performance not only by selecting the best employees, but also by influencing employees´ motivation and attitudes and develop positive mediating factors including improved productivity, positive social outcomes and reduced turnover (Kramar, 2014; Takeuchi et al., 2009).

Although work systems should serve the organization well while equally serving employees, there is a growing body of research questioning HPWS being mutually beneficial (e.g. (Godard, 2004; Kramar, 2014). Thus, Kramar (2014) argues that HPWS fails to adequately address the influence of different stakeholders´ perceptions. Keeble-Ramsay and Armitage (2015) point out that HPWS only has limited consideration for the complexity of inter-organizational relationships and often results in giving precedence to short-term financial targets reflecting shareholders´ interests (see also section 3.1.3).

While some researchers investigate the conditions under which HPWS can prove mutually beneficial (e.g. (Boxall, 2013) others claim that more sustainable and holistic approaches need to be adopted. Thus, during the last decade an approach labelled "sustainable Human Resources Management" (sustainable HRM) has emerged. Kramar (2014) identifies three distinct strands of literature concerning sustainable HRM. The first group, known as "Capability Reproduction" emphasizes internal and external outcomes and the creation of "sustainable competitive advantage". A second group focuses on broader performance outcomes such as environmental and social outcomes whereas the third group ("Connections") investigates the relationships between management practices and broader organizational outcomes. Particularly the first group has already produced a substantial body of literature. (Kramar, 2014). Thus, they identified HR practices that contribute to internal outcomes including employee engagement, job satisfaction and positive psychological contract. Among those practices are dialogical communication with employees, work roles and performance evaluation which focus on building on employee strengths and facilitating performance (Wells, 2011). Whilst some of the identified practices may resemble HPWP rather than new concepts, sustainable HRM writers point out that the focus is on long term, durable outcomes and on sustaining, developing and reproducing an organization´s human and social resource base (Kramar, 2014).

3.1.3 Critical evaluation of HRM

The call for alternative approaches in HRM clearly shows that currently there is a major stream of research taking a critical view on the role of HRM and related functions like Human Resources Development (HRD). Thus, Thompson (2011) warned that the construction of a special domain "critical HRM" would be "disastrous", but at the same

time he points out that “HRM is in trouble” because its core claims and professional self-image have been neglected. Considering HRM´s core claims, Ardichivili (2013) states that HRM has been criticized for removing itself from its original concern for the well-being of individuals and disentangling from its roots in the humanistic social science. Back in 2004, Godard found that the “high-performance paradigm” tends to have rather negative effects for employees. While indicating that HPW can have positive social and psychological implications for employees and yield positive organizational outcomes if adopted under the right conditions, he claims that mostly the paradigm does only yield marginal performance gains and may at least have ambiguous impact on workers (Godard, 2004). Since then, there have been several studies that report on a failed uptake of HPW and on difficult working conditions in Western organizations post-2000. In a comprehensive study, Keeble-Ramsay and Armitage (2010), for instance, investigated the perceptions of the shift towards HPW among more than 100 HR professionals in the UK. They found that most HR professionals perceived a move towards HPW practices being adopted by other organizations rather than in their own working environment. Hence, the authors conclude that there is a gap between organization´s aspirations and employees´ perception of HPW (Keeble-Ramsay and Armitage, 2010).

One of the major points of criticism deals with the question how HRM and HR practices were employed during, or may even have contributed to, the recent global financial crisis (GFC). MacKenzie et al. (2012) argue that by shifting the focus from operational HRD to strategic HRD, by advocating short term financial targets and “performance horizons” and by employing incentive structures that were devoid from any long-term social impact considerations, HRM has contributed to the GFC. Thompson (2011) claims that - due to a shift from managerial to financial capitalism - even organizations that develop track records of high performance and commitment were finding it harder to sustain to stable conditions that are a prerequisite for mutual benefit. This is also due to a disconnection of HR managers, who were tied to financialised practices and measures and therefore lacked the capacity to enabling stable conditions and to engage in people-focused management (P. Thompson, 2011). In consequence, many authors conclude that during the GFC, HRM abandoned one of its primary responsibilities: to maintain employees´ well-being and level of motivation and engagement during difficult periods. Rather, the majority of Western organizations reacted to the economic pressure by adopting control-oriented strategies and by applying a series of harsh practices including wage adjustments (cutting salaries, fringe benefits) or workforce adjustment (downsizing, layoffs) (Psychogois et al., 2016). Thus, the GFC in many countries led to a worsening of working conditions such as a steady increase in work intensification, long working

hours and a reduced commitment of employees left within these companies (Kemble and Keeble-Ramsay, 2014).

Another key issue when it comes to HRM criticism is the missing acknowledgement of the ambiguous position of HR managers, which was particularly visible during the GFC. Thus, HR professionals have to demonstrate that they are contributing to the financial outcomes and "adding value" to the organization. On the other hand, they have a commitment to serving the needs and well-being of the employees. In consequence, they need to take into account a variety of stakeholder interests. Particularly in the banking/financial sector, HR professionals have been criticised of putting the interests of the organization first and those of employees and society second (Kramar, 2014; MacKenzie et al., 2012). MacKenzie et al. (2012) point out that in the light of a "financialized culture" many HR practitioners may have found themselves on a back foot to senior management co-workers and may have had no choice but to remain silent if they wanted to save their own job. In the UK, this reluctance to speak up has left HRM in a "silenced" status prohibiting it from making valuable contributions to the development and well-being of the workforce (Keeble-Ramsay and Armitage, 2015).

In other European economies the impact of the GFC and the critical role of HRM may have not been this intense. As Godard (2004) observed: In co-ordinated market economies (e.g. Germany) the limitations of HPW are lower than in liberal market economies (e.g. the UK) since in co-ordinated market economies workers tend to have more co-decision and representation rights. Further, protection rights against layoff are stronger. Thus, during the GFC, German organizations issued a specific labour market policy – the "Kurzarbeit" – that allowed flexible short-time working and compensated employees working fewer hours. In many cases, "Kurzarbeit" approaches were supported by the Federal governments. This kept people in employment over the crisis and prevented a long recession (Psychogois et al., 2016).

However, regardless of the differences among European economies during the GFC and the different level of implications, the preceding analysis has shown that HRM has been subject to major criticism over the last decade and that the criticised issues are reasonable. As a consequence, many authors claim to develop bundles of HR practices that support organizations during crisis, including a constant dialog with employees about business planning and training to develop a broad skill set (Kramar, 2014; Psychogois et al., 2016). To resolve the crisis of "silenced" HRM, alternative approaches that take a critical perspective and facilitate dialog and development should be adopted (Keeble-Ramsay and Armitage, 2015). In this respect, the emerge of sustainable HRM (see section 3.1.2) seems to be a positive development.

3.2 Performance Management as a domain of HRM

Although a complex HRM system often consists of a bundle of intertwining HR practices, it becomes evident that performance management is a core HR function that organizations need to attend to irrespective of their specific HR strategic focus (Albrecht et al., 2015). In order to reflect relevant issues on managing performance, the evolution of performance management is discussed in this section. In addition, to identify the varying definitions of terms and to determine a consistent use of terms throughout this thesis, the key terms “performance management”, “performance measurement”, “performance appraisal” and “performance evaluation” are analysed more closely. Subsequently, approaches to manage performance on organizational and individual level are discussed and the impact of PM practices on performance is reflected.

3.2.1 Defining Performance

The term “performance” is used in a variety of meanings. For instance, it is used in sports as well as in technological contexts. In a rather biological context Darwin used the term to describe his ideas about tactics that species adapt to survive in a complex and uncertain environment. In this thesis, the term will be used in the economic, or more precisely, in the organizational context. Again, in the organizational context, there is no consensual definition or a commonly use of the term. This is mainly due to the multidimensional character of the performance concept. Thus, “performance” is often defined in financial terms (e.g. market value, profitability), operations (efficiency, number of outputs), marketing (number of customers, customer satisfaction) or others (Verweir, 2006). Moreover, the complexity of the term “performance” is reflected by a variety of definitions:

Definition	Source
"Performance is defined in terms of value that an organization creates using its productive assets in comparison with the value that the owners of these assets expect to obtain."	(Barney 1997)
"From a marketing perspective, organizations achieve their targets by satisfying their customers with greater efficiency and effectiveness than their competitors. Hence, in a business context, performance can be defined as the efficiency and effectiveness of purposeful action."	(Neely and Mills, 2002)
"Performance is defined as the record of outcomes produced on a specified job function or activity during a specified time period."	(Bernardin and Wiatrowski, 2013)

Table 2: exemplary definitions of performance

Most researchers concur that performance is strongly linked to efficiency and the ratio of output to input. Thus, performance is sometimes confused with productivity. Thomas and Baron (1994) argue that many people who claim to be discussing productivity are actually looking at the more general issue of performance. Tangen (2004) states: *„(…) the confusion surrounding productivity will be even more complicated with a too broad view of productivity"*. He suggests that *"performance, not productivity, should be used as an umbrella term for all concepts focus on the organization´s performance and its activities concerning this matter"*. Hence, "performance management (PM)" can be defined as *"a systematic process for improving organizational performance"* (Armstrong, 2018). Those will also be the definitions that will be adapted within this thesis.

The issue that an organization´s performance should be monitored and measured on a regular basis in order to achieve high performance standards has been equally taken up by practitioners and researchers (Cocca and Alberti, 2010). As a result, research into how to measure and manage performance accurately has become an increasingly popular field over the last two decades (see chapter 3.2.2). In this respect, Neely (2005) concluded that this area of research "has relatively little consensus about its core theoretical foundations". Folan and Browne (2005) argue, that this is due to the multidisciplinary character of this field of research. Therefore, it is not "owned" by academics of any particular discipline, which hinders development and often results in

performance management information that is duplicated or contradictory in nature. Nudurupati et al. (2011) report that there are several definitions, with no consensus between them. They conclude that this is due to the obscurity in this research field. Therefore, it can be noted that the understanding of terms in performance management literature is widely diverse. In order to transfer performance management research into a cohesive body of knowledge, Folan and Brown (2005) claim an examination of performance measurement within the distinct concept of performance management. They indicate that both performance measurement and performance management follow one another in an iterative process and in doing so create the context for their existence (Folan and Browne, 2005).

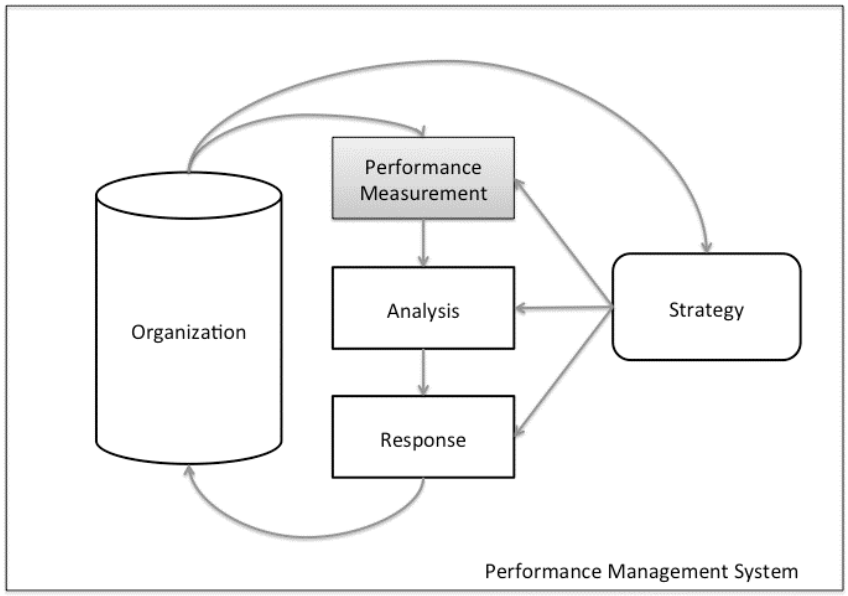


Figure 5: Schematic representation of a performance management system (Folan and Browne, 2005)

Thus, performance measurement is an essential part of performance management. In order to provide performance-based data that can easily be converted into performance management activities, performance measurement should be integrated into a comprehensive performance measurement system (PMS) (Star et al., 2016). The aims, requirements, stakeholders and design of a PMS are illustrated in chapter 4.

3.2.2 The Evolution of Performance Management

Historically, the role of performance management evolved over time. At the beginning of the twentieth century, Frederick Taylor was the first to introduce notions of performance and efficiency. He believed that it was management´s responsibility to devise the best method of performing work. Thus, he proposed a mechanistic management approach that became known as “scientific management”. It was based on the analysis of existing

work methods through observation and measurement (Martinez et al., 2007). This approach was seen to legitimize management as a control agent. It was enhanced by many other researchers and shaped the work of highly successful managers like Henry Ford or theorists like Henri Fayol. While Ford applied the approach to develop early mass production systems, Fayol derived key managerial functions including planning, organizing and control (Hailey and Sorgenfrei, 2004). At individual level, the scientific management approach laid its focus on the individual output of each worker. Accordingly, on organizational level, managers mostly looked at accounting-based performance measures (Yadav and Sagar, 2013). Thus, performance management was primary volume and cost centered.

During the 1950s and 1960s the perception of performance management shifted. On the one hand, there was criticism on the scientific management approach that evolved from the human relations movement. The supporters of the human relations movement argued that social factors of work were at least as important as the technical ones advocated by the scientific management approach. On the other hand, in the era of low unemployment, management´s focus shifted to concerns on how to attract and retain employees. Thus, a greater emphasis was put on the improvement of working conditions and on the assessment of performance-related inputs factors like competencies (Houldsworth and Jirasinghe, 2006; Martinez et al., 2007).

During the 1970s, the western style management practices were extremely challenged by manufacturers across the world, notably Japanese ones. In this respect, many managers questioned the use of traditional performance measures, since Japanese manufactured goods not only were of greater variety but also of better quality. Thus, they called for better measures that could assess why some factories were better than others. Subsequently, Total Quality Management (TQM) entered the performance management scene in the 1980s and with this, new performance measures like defect rates, response time or delivery commitments came into picture. From the quality management paradigm, different approaches including Six Sigma or Lean Enterprises emerged. Despite the recognition of the quality approaches, the organization´s accounting systems merely included to financial information (Nudurupati et al., 2011).

By the 1990s, quality approaches were well-established and many researchers started to criticize using financial figures only, indicating that qualitative measures like customer satisfaction, competition indexes or innovation should be integrated into performance management (Hailey and Sorgenfrei, 2004). By introducing the Balanced Scorecard, Kaplan and Norton (1992) brought kind of a “performance management revolution” proving that complementary to financial figures operational and strategical measures can

be included into one PM framework (Yadav and Sagar, 2013) (see also chapter 3.2.4). Futher, the BSC introduced the involvement of different stakeholders including employees, customers or suppliers. Since then, several integrated PM systems and frameworks emerged and with them the number of performance measures rose (Nudurupati et al., 2011).

In the mid to late 90s, simultaneous to the excessive use of performance measures, the development reached its peak with one book on this subject being published at a rate of one every two weeks in the U.S. alone. Further, between 1994 and 1996, more than 3,600 articles on performance measurement were published (Folan and Browne, 2005; Star et al., 2016). The increasing range of measures led to what Neely and Austin (2000) described as the “first measurement crisis” with organizations becoming more concerned about measuring than actually managing performance.

During the last two decades, the focus of performance management again has shifted from looking across organizations functionally to a rather horizontally (process-oriented) view. Management techniques like Business Process Reengineering (BPR) promoted this view by focusing in particular on processes that crossed departmental boundaries. Both, TQM and BPR shifted the focus of performance management from “producing” to “improvement” (Martinez et al., 2007). In 2006, Davenport noted in times where organizations offer similar products and services by using the same technologies, business processes are among the last remaining points of differentiation. He introduced “Business Analytics” as an organization´s ability to collect, analyze and act on data (Davenport, 2006). Since then, business analytics is an emerging field that has also gained momentum in performance management (Bronzo et al., 2013). The aim of business analytics in the domain of performance management is to supplement this area with sophisticated and analytical decision-making tools, which lie beyond the domain of traditional performance management including mathematics, statistics, econometrics as well as tools for data gathering and analysis. In consequence, business analytics shall provide inside into business dynamics and their related performance in order to use analytical indicators to predict performance (Schläfke et al., 2012).

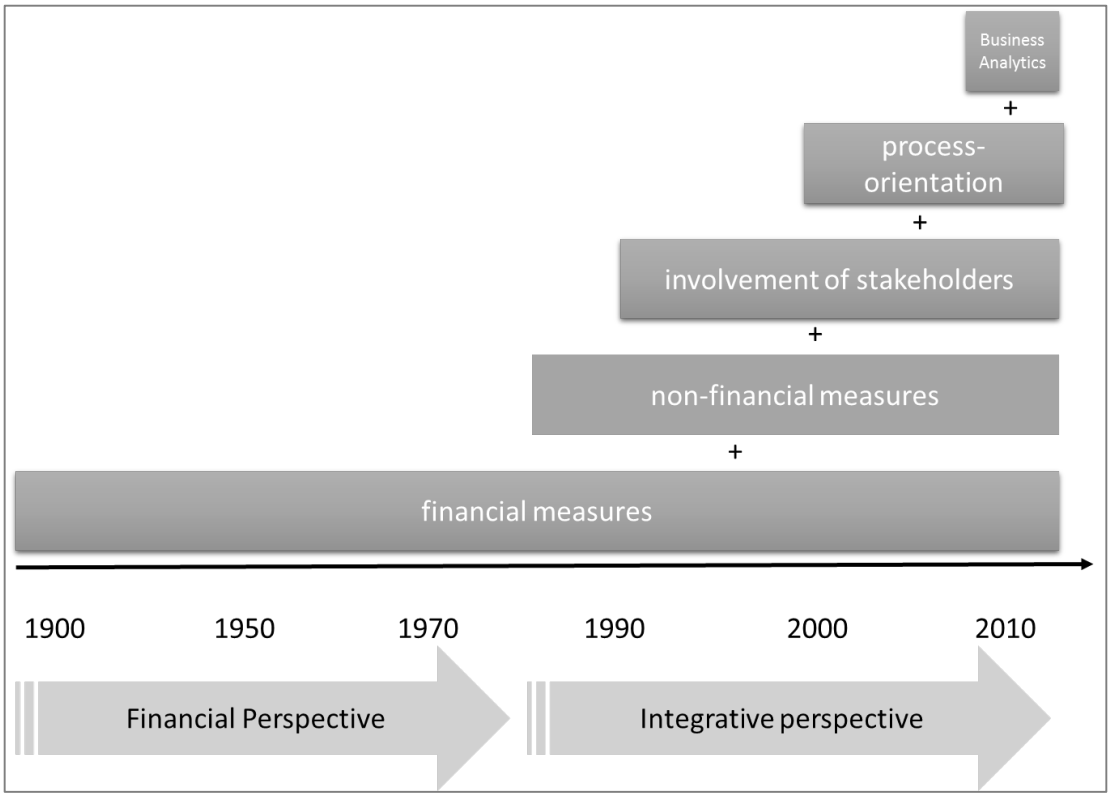


Figure 6: The Evolution of Performance Management

Summarizing, there are three general trends in the evolution of performance management:

1) The broadening of analysis: From a mere financial perspective to an integrated perspective
2) The broadening of performance measures: from cost and output to quality and input.
3) The broadening of the scope: from internal focus towards the inclusion of external focus (stakeholders) and the application of business analytics.

3.2.3 Managing individual performance

In spite of some exceptions (e.g. Becker et al. (2001)), most authors assign the application of performance measurement and performance management on organizational or unit level (Micheli and Kennerley, 2005; Sebald and Jacob, 2015). However, since performance can also be addressed on individual level, there needs to be another term referring to individual performance management.

A term that is often linked to the management of employee performance is the term “performance appraisal”, which are sometimes also labelled as “performance evaluation” or “performance review”. In general aim to ensure that employees´ performance contributes to the organization´s targets by evaluation employees´ strengths and weaknesses in a continuous manner (Aguinis et al., 2011; Chartered Institute of Personnel and Development, 2017). Since first applied during the Industrial Revolution, performance appraisals were rather control-oriented emphasizing individual accountability and supporting transfer or layoff decisions. A severe shortage in skilled labour caused a shift in priorities in many organizations during the late 1950s. The emphasis now was on developing employees´ strengths. During the 1970, with inflation rates rising, merit pay systems were the central focus of HRM. Thus, performance appraisals became more administrative again in order to identify top performers. By the early 2000s, most organizations applied performance appraisals merely to allocate pay and rewards and to hold employees accountable (Cappelli and Tavis, 2016).

At the same time, negative attitudes towards appraisals rose by managers and employees. Many supervisors struggled with distinguishing good performers from bad, often lacking time and methods to adequately evaluate them. Also, employees expressed concern that their performance has not been fairly rated. Many studies also indicated that, although there was proof that both organizations and employees can benefit from using performance appraisal, they often remain under-fulfilled when they come to their practice (Kim and Holzer, 2016). In a comprehensive study Fletcher (1997) found that most UK organizations were dissatisfied with their appraisal process as it often failed to give valid performance ratings and did not succeed in developing skills and motivating employees. Further, Kellough and Nigro (2002) found that performance appraisal even may result in perceived stress, demotivation or even burnout.

Since dissatisfaction with the appraisal rose, some organizations – mainly from the high tech industries - abandoned traditional performance appraisals by the beginning of the 2010s, replacing them with more frequent and informal exchanges (Chartered Institute of Personnel and Development, 2016a). Adobe, for example, launched the “Agile Manifesto” breaking projects into “sprints” that were followed immediately by a debriefing session. Thus, they included constant feedback and assessment into performance evaluation (Cappelli and Tavis, 2016). Also Netflix abandoned all formal reviews, replacing them with rather informal conversations (Chartered Institute of Personnel and Development, 2016a).

In consequence, some researchers proclaimed a “Performance Management Revolution” (Cappelli and Tavis, 2016) or “reinvented Performance Management”

(Chartered Institute of Personnel and Development, 2016a). While it seems reasonable that approaches to managing employee performance change over time, there is currently no evidence that those "new" approaches actually improve performance. In this context, the Chartered Institute of Personnel and Development (2016a) provided a comprehensive analysis, reviewing the body of evidence-based research on the causes and effects of performance appraisals. In conclusion, they found that the performance appraisal is an essential part within the performance management process and that "getting the performance appraisal right makes the difference between a positive impact, a negative impact and no impact". In order to "get the appraisal right" the CIPD (2016a) identified several key points:

- It is not the appraisal process or method per se that is important, but the employees´ reaction to it.
- Despite the varying emphasis given to them over the past decades, both administrative and developmental purposes are important to the appraisal. However, they should be discussed in separate meetings.
- The aim of the performance appraisal method that is in place needs to be clear with employees at all levels.
- Fairness is essential to the perception of the appraisal. It is by far the most consistently researched aspect.

All of the mentioned issues will be discussed in more detail in chapter 3.2.9. In the light of its partially negative connotation to employees and managers, the term "performance evaluation" instead of "performance appraisal" will be applied throughout this thesis.

3.2.4 Approaches for managing performance

Considering the development of HRM, some researchers claim that it has developed on two parallel tracks: a micro perspective with focus on individual development, employees´ well-being and retention and a macro perspective rather focusing on the organization´s HR strategies and practices (Takeuchi et al., 2009). To a certain degree this is also true for performance management. In research and in practice, the term "performance management" relates to the performance of individuals as well as to the performance of the whole organization. On individual level it is often aimed at evaluating and improving employees´ individual performance. Therefore, it mostly is initiated and executed by the organization´s HR department. On organizational level, performance management generally is understood as the financial and organizational management

of the organization´s performance, which is usually a coordinated by the controlling department (Sebald and Jacob, 2015).

As a result, the vast majority of PM approaches focuses either organizational or on individual level, whereas some approaches (e.g. the Balanced Scorecard or the EFQM framework) blur the division and are applied on both levels (Pleier, 2008). However, there is a growing cognizance among researchers and practitioners that there is no single PM process or approach that guarantees sustainable success and competitive advantage. Rather performance management on individual and organizational level should be interdependent in order to align corporate strategies with operational targets as well as organizational performance with individual development (Pun and White, 2005; Sebald and Jacob, 2015). In consequence, a holistic Performance Evaluation System (PES) could be established integrating the organizational and the individual perspective.

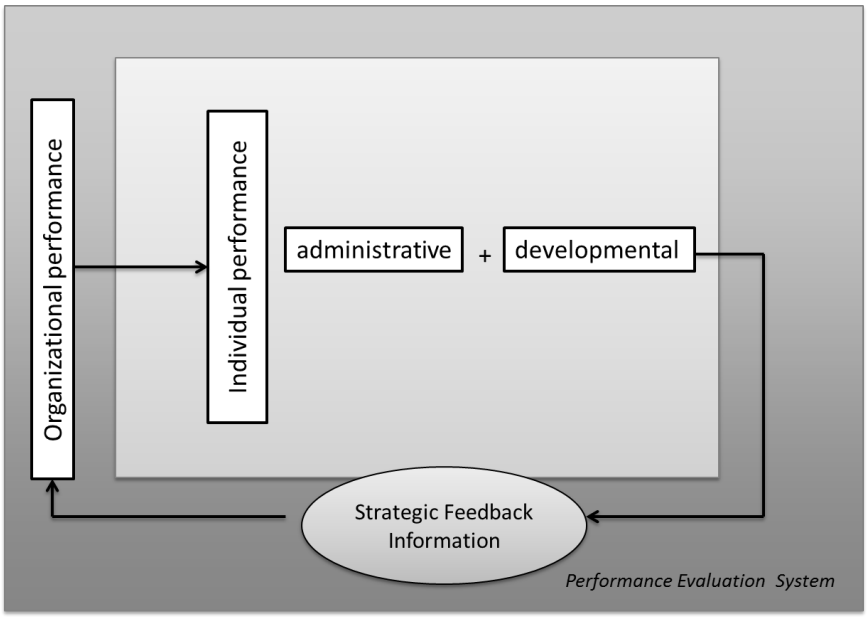


Figure 7: Integrated Performance Evaluation System

In the following chapters, a comprehensive overview over PM approaches and frameworks on organizational and on individual level is provided.

3.2.5 PM Approaches on organizational level

For managing organizational various authors have developed PM frameworks and approaches supporting the organization to translate strategy into results and controlling their business processes. The advent of PM approaches can be dated back to the 1940s

where PM was generally undertaken to broaden an organization´s accounting system. Since then the evolution and development of PM approaches led to an immense growth of various approaches and frameworks (Star et al., 2016). This growth reached its peak the 1990s which brought a revolutionary development and with this drastic changes in the way organizations measured performance (Yadav and Sagar, 2013). Since some of the most frequently used PM approaches evolved during the 1990s, this section starts by providing an overview of the most popular and most cited approaches and frameworks of this era. However, more recent studies report that research related to PM approaches meanwhile has reached a second generation with a focus on the exploration of more proactive and more holistic approaches (Taticchi et al., 2010; Yadav and Sagar, 2013). Thus, more recent approaches from the era beginning 2001 are also discussed.

3.2.5.1 First generation approaches (1990 – 2000)

The results and determinants framework

Many of the frameworks emerging in the 1990s tried to offer integrated solutions linking strategy to operations and using balanced measures as well as non-financial performance indicators. One of the first approaches to offer an integrated solution is the results and determinants framework (RDF) which was proposed by Fitzgerald et al. (1991) who examined performance in for profit services (Pun and White, 2005). It is based on the premise that there are two distinct types of performance dimensions: those that focus on results and those that focus on the determinants of results. It emphasizes that results obtained are a function of past performance and therefore are lagging indicators. Again, results are due to specific determinants, which are leading indicators (Neely et al., 2000). The core elements of the RDF are illustrated below.

Results	Financial Performance
	Competitivness
Determinants	Quality
	Flexibility
	Resource utilisation
	Innovation

Figure 8: Measurement Framework Fitzgerald et al. (1991)

Whilst the framework incorporates two perspectives and non-financial measures, the measures do not vary across different types of services or organizations. Thus, it fails to integrate other leading indicators or other non-financial figures (Yadav and Sagar, 2013).

The Performance Pyramid

To integrate different hierarchical levels and to link performance across those levels, the Performance Pyramid (PP) was introduced by Lynch and Cross (1991). The development of the framework starts with defining corporate vision cascading them down into business unit objectives and operational objectives. Therefore, the framework consists of four levels of objectives and measures with corporate vision and strategy on top and department level as a fourth level. The operational measures at the bottom are key to achieving higher-level results (Pun and White, 2005).

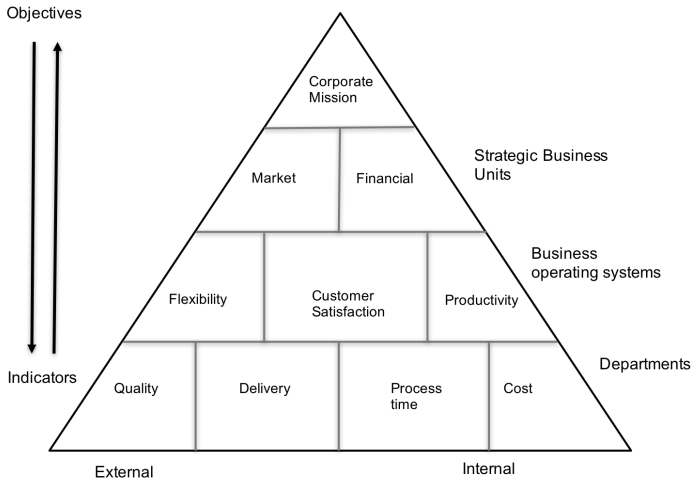


Figure 9: Performance Pyramid, (modified from Pun and White, 2005)

Causal relationships of objectives or measures are visualized by so called “building blocks of success”. Performance loops that are integrated at each level shall ensure constant adjustment to change (Pleier, 2008). Thus, the PP´s strengths lie in bridging the gap between top level and operation floor by integrating corporate objectives with operational measures. However, one major point of criticism is that PP fails to provide a mechanism to identify key performance indicators (Yadav and Sagar, 2013). Furthermore, its focus is mainly on two stakeholders: management (internal focus) and customers (external focus). Employees are recognized as contributers to organizational success, but their specific interests are not integrated (Pleier, 2008).

The Balanced Scorecard

One of the most popular PM approaches – and listed as one of the 75 most influential ideas of the twentieth century by *Harvard Business Review* (Yadav and Sagar, 2013) - is the "Balanced Scorecard" (BSC) introduced by Kaplan and Norton in 1992. The Balanced Scorecard focuses on four business perspectives addressing four basic questions (Kaplan and Norton, 1992):

1. Financial perspective: How do we look at shareholders?
2. Customer perspective: How do customers see us?
3. Internal business perspective: What must we excel at?
4. Innovation and learning perspective: How can we continue to improve and create value?

The "balanced" set of performance dimensions brings together seemingly disparate elements of an organization´s strategic agenda, incorporating financial and non-financial figures and provides users with a single report (Kaplan and Norton, 1992; Yadav and Sagar, 2013). Whilst the perspectives are mostly predetermined, the associated performance measures are explicitly arbitrary (Folan and Browne, 2005). Thus, BSC aligns strategy with targets on strategy, department and individual level.

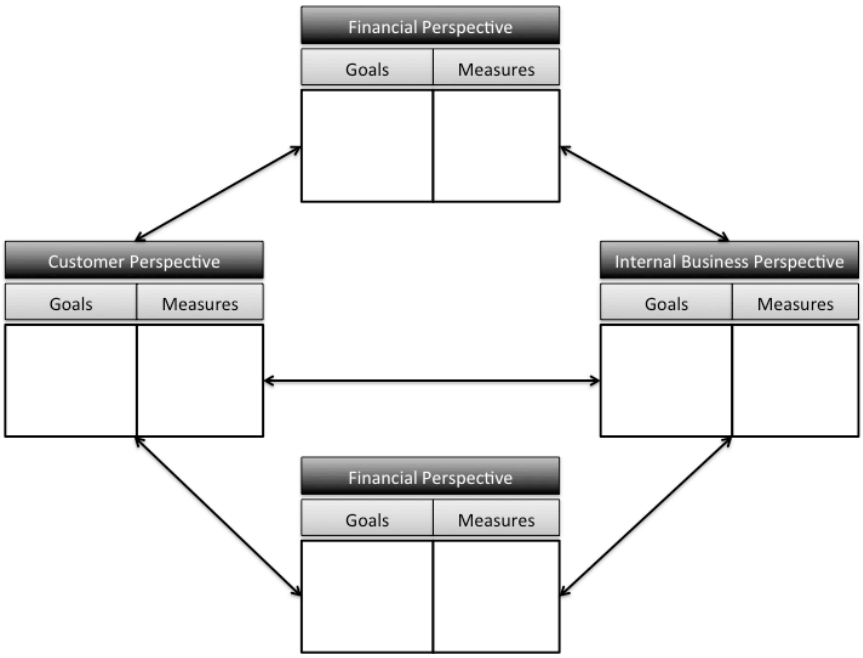


Figure 10: The Balanced Scorecard (Kaplan, Norton 1992)

In contrast to the Performance Pyramid, the BSC helps management to identify and concentrate on indicators that have the greatest effect on organizational success (Pun and White, 2005). As a multi-measure and multi-stakeholder approach the BSC acknowledges three key stakeholders: shareholders, customers and employees. In contrast to other approaches – the Performance Pyramid, for instance - regarding shareholder, customer and employee interests as distinct and unrelated, the BSC sees all three as being bound together in the organisational 'process of "value creation" (Shields et al., 2015).

However, despite its widespread use and strengths, BSC has several weak points. For instance, some authors claim that BSC does not provide accurate and fast feedback or benchmark information and does not involve key users in the development process (Star et al., 2016; Tan et al., 2017). Further, the approach is often criticized for its static nature lacking loop learning and for implementation failure (Yadav and Sagar, 2013). Also, some authors claim that there are inconsistencies in the way the balanced scorecard is researched. In particular, the identification of cause-and-effect relations by doing empirical research is found to be challenging. In this context, the assumption of a linear causal relationship between people, processes, customers and profits is questioned by researches and practitioners (Shields et al., 2015).

3.2.5.2 Second generation approaches (from 2001)

The Performance Prism

Among the first of the "second generation approaches" is the Performance Prism (PPR), which was introduced in 2001 by Neely et al. (2001). According to Neely, the PPR considers some of the failings of its predecessors and therefore, for instance, encompasses a broader, more holistic view than BSC and other first generation approaches. Thus, the PPR is a stakeholder-focused approach including stakeholders like employees, suppliers, local authorities and other interested groups (Star et al., 2016). The PPR is built of five facets. Since the approach´s philosophy supports a reciprocal relationship between organization and stakeholder, the facet "stakeholders" is the first facet of the prism imposing the questions "Who are our stakeholders?" and "How can we meet their needs?". The other facets are:

- Strategies: "What are the right strategies to deliver value to our stakeholders?"
- Processes: "What processes are required for implementing our strategies?"

- Capabilities: “What capabilities (combination of people, practices and technology) enable execution of the organizations Business Processes?”
- Stakeholder Contribution: “What is the stakeholders´ contribution back to the organization?”

Especially the last facet is a novel way to define the stakeholder relationship. To develop a PPR performance indicators and measures that are oriented towards stakeholder´s needs and expectations are to be derived (Neely et al., 2001; Star et al., 2016).

With its explicit focus on multiple stakeholders and the reciprocal relationship between stakeholders and organization, the PPR offers two essential enhancements to other PM approaches (Pleier, 2008). In terms of disadvantages, the PPR shows a lack of an ongoing review process. This weakness meanwhile was addressed by several authors (e.g. Najmi (2012)), who created revised frameworks (Star et al., 2016).

Updates on the BSC Approach

The second generation of PM approaches also brought many developments updating the traditional BSC. Among the most discussed in literature is the Comparative Business Scorecard (or Kanji´s BSC) by Kanji and Sá (2002) who also emphasize the role of the stakeholders by defining stakeholder value and satisfaction as two of five critical dimensions for business success. Furthermore, they integrated a cycle of continuous improvement, which is claimed as one of the weak points of the traditional BSC (Pun and White, 2005).

Other developments like the holistic scorecard (Sureshchandar and Leisten, 2005) enhanced the BSC focus by integrating more dimensions, for instance an employee perspective or a social perspective. To support complex decisions, Barnabè (2011) applied the system dynamics methodology to modify the traditional BSC integrating feedback loops and dynamic strategy maps. To draw causal representation of key performance indicators and to simulate and quantify their impact, Chytas et al. (2011) applied fuzzy logic for a more proactive view of the BSC. Thus, they integrated simulation techniques to test the feasibility of policies by showcasing future results (Yadav and Sagar, 2013). In order to support business sustainability and performing HR sustainability, Maurer and Müller-Camen (2016) proposed the sustainable HRM scorecard by defining recruitment, compensation and employee well-being as essential dimensions of sustainable HRM.

Dashboards

In the narrow sense, dashboards are not approaches for measuring organizational performance. Rather, they are reporting tools for giving users better access to crucial information in a way that prevents information overload and provides them with performance data in timely and functional display. Dashboards therefore rely on visual techniques such as the use of colours, dials, traffic light icons or graphs. Traffic light icons, for example, can be used to provide information if performance objectives and actual performance are close or if there is a discrepancy (Bremser and Wagner, 2013).

The use of dashboards to monitor strategy implementation at various organizational levels is an increasing trend since the 2010s. One of their major advantages lies in their strength in filtering critical information and providing it in an appealing manner. Further, the development and implementation only needs little investment at the beginning. However, they need to be examined and updated frequently to guarantee timely information and prompt adjustments if data requirements change (Star et al., 2016).

Business Analytics

As the need to filter and edit performance information with the help of dashboards clearly shows, organizations gather a large amount of data but hardly manage to analyse all of it. With a seminal article in 2006, Davenport promoted the use of analytics to gain insights into business dynamics and exploit performance information to identify and actively control key performance drivers (Davenport, 2006). The work of Davenport rapidly gained repercussion among both practitioners and researchers (Bronzo et al., 2013). In 1995, Neely, Gregory and Platts already claimed that the diverse field of performance management with its focus on different aspects of a large number of measuring systems is a persistent problem in performance management research. The approach of Business Analytics acts on this problem by supplementing traditional management approaches with more complex and sophisticated mathematical, statistical or econometric models. By the inclusion of analytical tools, PM should become more forward-oriented and helpful to identify and forecast the dynamics that have an impact on organizational success. Thus, Business Analytics could potentially increase PM´s effectiveness by discovering new or hidden business dynamics at strategic or organizational level or prove causal relationships between input and output factors. This way, Business Analytics can help to identify potential strategies and support decision-making in complex situations (Schläfke et al., 2012). For the use of Business Analytics in the PM context, Schläfke et al. (2012) propose the term “Performance Management Analytics (PMA).

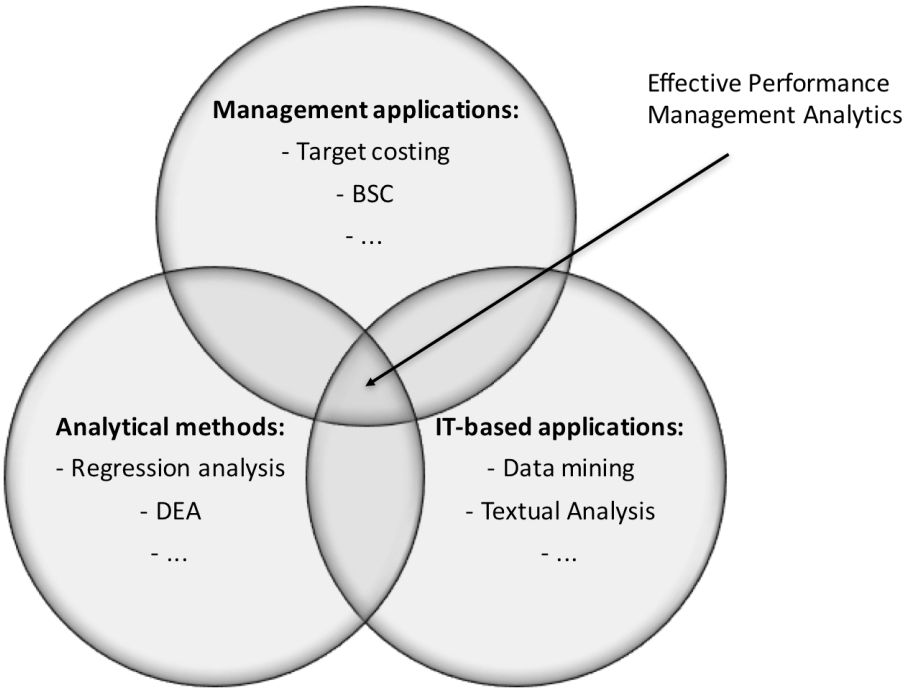


Figure 11: Application of PMAs (modified from Schläfke et al. (2012))

According to Schläfke et al. (2012), effective PMA approaches emerge at the intersection of traditional management applications, IT-based applications and analytical methods. In this sense - as an analytical method - DEA could be part of a PMA (the integration of DEA in a PMS and its delimitation from PM approaches on individual level is discussed in section 3.2.7).

Despite its apparent advantages over traditional approaches, the implementation of PMA places high demands on organizations. First, organizations need to engage into technological tasks including making data available in warehouses, selecting and implementing analytical software and assembling the hardware environment (Davenport, 2006). Further, managers and employees may need to acquire new skills (mathematical, statistical) to be able to employ PMA. Finally, the process of gathering data can be a problem since some drivers of performance are hard to measure (Schläfke et al., 2012).

3.2.6 PM Approaches on individual level

3.2.6.1 Classification of approaches

Although the value of traditional performance appraisals or evaluations has been increasingly questioned in recent years, most researchers agree that they remain a crucial element of performance management (Chartered Institute of Personnel and Development, 2017).

To evaluate performance on individual level, there are a variety of approaches known to literature and practice. In general, measurement approaches are horizontally classified into three categories: behaviour, results and trait approaches (Aguinis, 2009).

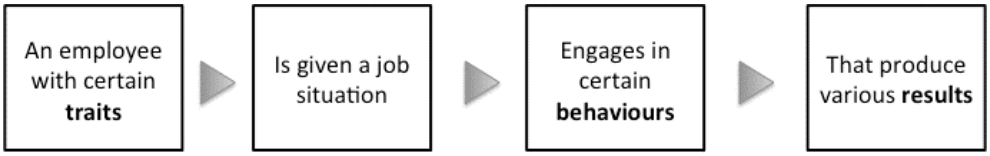


Figure 12: Job performance in context (modified from Aguinis, 2009)

Trait approaches do not consider behaviours or results, but focus on the individual´s traits that may include cognitive abilities like knowledge, skills and competencies. Strictly speaking, traits are not synonymic to performance. Rather, they are the precondition for employee's potential to perform (Shields et al., 2015). Since the link between competence and performance is not yet proven, many organizations remain sceptical and only a minority of organisations have thus far adopted competence-based approaches as the centrepiece of their performance management practice (Shields et al., 2015).

Behaviour approaches focus on factors expressing how an employee does the job. In this context, behaviours are seen as activities that transform inputs into outputs, including, work effort and other behaviours (Shields et al., 2015). Outcomes of performance are usually not considered. Thus, behaviour approaches can be interpreted as a rather input-oriented. (Manoharan et al., 2009). These approaches are particularly appropriate when the link between behaviours and outcome is not obvious or outcomes occur in the distant future. Also, they should be appropriate to non-routine work situations with minimal supervision, and with numerous pathways to get to the same result. This is often the case in knowledge-based work organisations or professional services (Shields et al., 2015).

Results approaches, on the other hand, focus on outcomes and results produced by the employee and therefore can be interpreted as rather output-oriented approaches. Since output measures often are of quantitative nature, results approaches are often seen as more cost-effective and more objective than behaviour-approaches. They particularly apply to work situations where individual outcomes can be accurately specified, quantified and measured (Aguinis, 2009; Shields et al., 2015) . Most organizations adopt behaviour or results approaches to assess performance. However, these two are not mutually exclusive. Some organizations use approaches that combine both behavioural

and results elements (Aguinis, 2009). To broaden the scope and maintain a consistent terminology throughout this thesis, behavioural and trait approaches will be referred to as “input-oriented” approaches and results approaches will be referred to as “output-oriented” approaches subsequently. Supplementary to horizontal classification, input-oriented approaches also can be grouped into two broad categories: comparative and absolute approaches. While comparative approaches focus on developing ranking of individuals within a given group of employees, absolute approaches seek to rate individual performance against a set of given criteria and to determine an absolute numerical rating for each employee (Shields et al., 2015).

The general classifications of PM approaches are illustrated below:

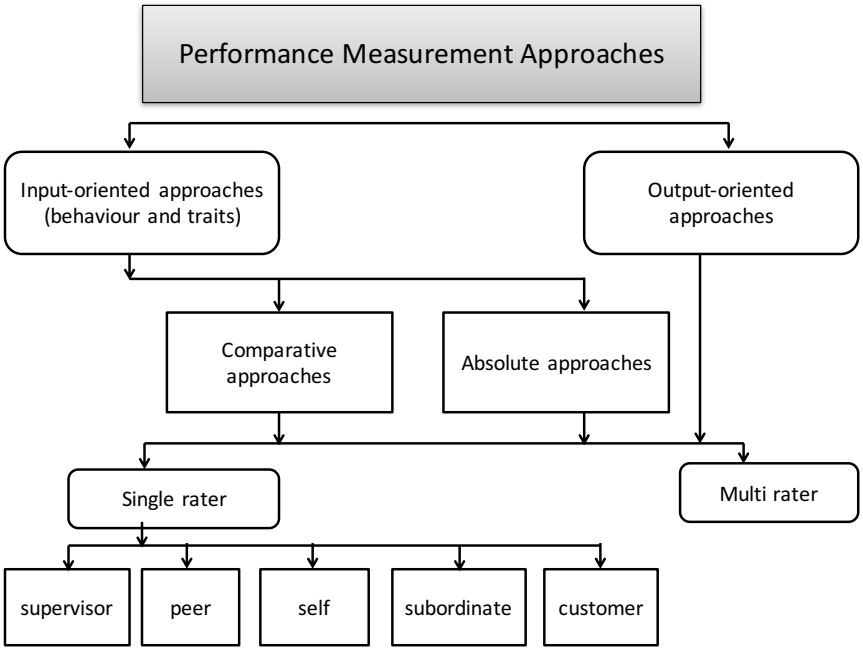


Figure 13: classification of performance measurement approaches (modified from Manhoran et al., 2009 and Shields et al. (2015))

From the variety of classifying characteristics result a variety of evaluation approaches that could be applied to measure and evaluate employee performance.

Input-oriented approaches		Output-oriented approaches
Absolute Approaches	**Comparative Approaches**	
Critical incidents appraisal	Straight ranking	Target Setting
Graphic rating scale	Paired comparison	BSC
Behavioural Anchored Rating scale (BARS)	Forced distribution	Visual Assessment
360-degree feedback		

Table 3: Performance Management approaches on individual level

In the following section, approaches that are among the most cited in literature and the most used in HR practice (Aguinis, 2009; Jafari et al., 2009; Shields et al., 2015) are discussed, pointing out some advantages and disadvantages of each.

3.2.6.2 Review of approaches

a) Input-oriented absolute Approaches

Critical Incidents Appraisal

The critical incidents appraisal is an approach that focuses on collecting direct observations on employees´ behaviours during different job situations. Thus, the supervisor gathers and writes down situations in which the employee exhibited behaviours that were either especially effective or especially ineffective to accomplish the job (Jafari et al., 2009). The resulting critical incidents report provides a starting point for evaluating performance. In a narrow sense, the critical incidents appraisal is not itself a behavioural rating method. It rather identifies performance criteria on which valid ratings may be based (Shields et al., 2015). One strength of this approach is its direct relation to the individual´s job behaviour rather than to vaguely defined traits (Aguinis, 2009). Further, it allows supervisors to give specific feedback on behavioural strengths and weaknesses. In addition, the use of specific diaries or forms is suitable to minimise the possibility of recency error (Shields et al., 2015).

One of the approach´s major weaknesses is its lack to attach a quantifying measure to the incidents and therefore administrative decisions are difficult. In addition, the gathering and reporting of critical incidents is very time consuming (Aguinis, 2009; Jafari et al., 2009). For supervisors, it may be challenging to subjectively select and weight incidents in terms of their relative importance to job performance. Thus, the approach is prone to both, unintentional and also intentional error (Shields et al., 2015)

Graphic Rating Scale

The graphic rating scale is a widely used qualitative technique for measuring employee performance. It consists of a list of categories where the employee's performance is rated against each one of these categories. The rating scale usually consist of three, five or up to seven grades of performance. For each criterion, the assessor should select the most appropriate grade of performance. Finally, the individual grades are aggregated arithmetically or intuitively to determine an overall performance score reflecting the individual level of performance. (Armstrong and Taylor, 2014; Shields et al., 2015). One major advantage of graphic rating scales is that they are easy to develop and to carry out. Further, the rating provides an absolute measure of performance differences between employees (Armstrong and Taylor, 2014). Since rating scales enable the evaluation of individual performance against a unique set of performance criteria, they stand to strengthen inter-assessor consistency and reliability (Shields et al., 2015).

Rating scales, however, are subject to some major points of criticism. Among the most cited weaknesses is the issue that rating scales have a lot of potential for subjectivity and bias. For instance, they are prone for unintentional rater bias like the Halo effect or anchoring effects. Managers also tend to give less favourable ratings if they received negative feedback themselves or are overly conscientious (Chartered Institute of Personnel and Development, 2017). Other authors meanwhile claim, that the rating of an employee´s total performance by a single score is a gross oversimplification (Armstrong and Taylor, 2014; Cappelli and Tavis, 2016). Further, an empirical study by Bol (2011) indicates that simple ratings often fail to differentiate performance among employees sufficiently. In a comprehensive empirical study among German financial institutions, Kampkötter and Sliwa (2011) found that on average, stronger differentiation has a substantial positive effect on performance.

Behaviourally Anchored Rating Scale (BARS)

The behaviourally anchored rating scale (BARS) is a graphic rating scale that uses critical incidents as anchors. It is developed in several stages. First, important dimensions of a job (category) are identified. Thus, specific incidents of effective and ineffective behaviours are identified and later grouped into broad behavioural categories. In a second step, critical incidents indicators that illustrate superior, average and poor performance for each dimension are determined and integrated into a rating scale, typically a 1-5, 1-7 or 1-10 scale, similar to a graphic rating scale (Aguinis, 2009).

By combining the critical incidents approach and the graphic rating scale, the BARS approach addresses some of the major shortcomings of both approaches. By providing detailed descriptions for both the behaviour itself and the grading scales by which behaviour is rated, the BARS seeks to overcome reliability problems that are inherent in graphic rating scales. Since the behavioural categories are predetermined, the BARS only looks at relevant job behaviour, not personal impressions, which may occur with the critical incidents technique. Due to clearly defined standards, it is able to provide feedback information for developmental purposes but also informs administrative decisions since it yields a total performance score (Shields et al., 2015).

Despite the BARS´s aptitude to address several shortcomings of other absolute approaches, there are also some drawbacks regarding this approach. First, like most behavioural approaches, it is prone to unintentional error and content invalidity since behaviourally anchored rating scales take no systematic account of the frequency with which particular types of behaviour are manifested. Hence, an assessor may evaluate an employee on the basis of just one observed incident ignoring the frequency with which

this behaviour is demonstrated. In addition, behavioural categories may facilitate the assessor´s recall of unrepresentative behaviours, which again may result in an unfairly low rating for a particular employee. Moreover, it is likely that assessors overlook the fact that employees should only be held responsible for performance indicators that are within their control by holding them fully accountable for observed behaviour (Shields et al., 2015).

360-degree feedback

The 360-degree feedback is a multi-source approach that involves evaluations from subordinates, peers and managers as well as self-assessment. Sometimes external feedback from customers is also included. Subsequently, the person receiving the feedback can compare her self-rating with the feedback provided by others.(Jafari et al., 2009). The rationale for gathering feedback from multiple sources is managers´ lack to oversee all performance aspects of the employees they manage in complex job situations or large organization. To conduct a 360 degree feedback, a range from eight to ten people (typically peers, senior colleagues, or customers) completes questionnaires or answers structures interviews to assess or describe one particular employee´s performance (Chartered Institute of Personnel and Development, 2018). In this context, the evolution of IT has caused a remarkable growth in the approach´s utilization, which allowed feedback data from multiple sources to be analysed and presented with increased accuracy and at a much larger volume (Bracken et al., 2016). Due to the involvement of multiple sources, on major advantage of the 360-degree feedback is that it reduces individual bias. In addition, it offers a more participative and inclusive process of reviewing employees' performance (Armstrong and Taylor, 2014).

On the other hand, the concept has several drawbacks, which led some authors to the conclusion that the 360 degree feedback has passed its peak (Bracken et al., 2016; Shields et al., 2015). Most critics claim that it merely replaces single-assessor subjectivity with multi-assessor subjectivity and assessor bias remains an ever-present possibility (Bracken et al., 2016; Shields et al., 2015). Particularly subordinates and colleagues may withhold negative ratings and/ or do not feel accountable for follow-through (Shields et al., 2015). Bracken et al. (2016) indicate that several organizations had problems with the implantation and the subsequent use of the concept. For instance, they found that employees often receive no assistance with interpretation. Managers on the other hand lack development and planning resources, follow-up accountability and tracking mechanisms. Also, Armstrong (2018) reports that several studies on the effectiveness of the approach showed mixed results. Whilst there are some rare success stories, most

studies showed that the feedback process had not been effective in changing performance and that in some cases, due to the mentioned issues, it has “done more harm than good” (Silverman et al., 2005).

a) Input-oriented Comparative approaches

Simple Rank Order

A simple rank order ranks the employees in an order from highest to lowest performer on the basis of comparative performance (Aguinis, 2009). Its appeal lies within its simplicity being easy to conduct and explain (Armstrong, 2018). Like other behavioural approaches, a simple ranking is highly subjective and prone to rating errors. There are some other weaknesses as well. Typically, simple ranking methods place their focus on whole-person judgements lacking specific performance criteria, which leaves no room for developmental analysis and therefore no feedback can be provided. Further, simple rankings hold no information on the relative difference between employees, which does not allow meaningful comparisons between employees or work groups ranked by different assessors (Armstrong, 2018). Also, they assume that all employees are comparable to one another. Another point of criticism is it may be difficult to rank the performance of average performers, who are typically the majority of the workforce. In addition, while in some cases rankings may be motivating to achieve a better rank, often they have serious negative effects on subsequent worker morale and task behaviour since they lack clear and understandable criteria (Shields et al., 2015)

Paired Comparison

A paired comparison makes explicit comparisons between all pairs of employees under review. Thus supervisors compare systematically each employee against the performance of all other employees with regard to standardized performance criteria Each employee is then ranked with regard to the number of criteria for where she performed better than the employee she was compared to (Aguinis, 2009).

By ranking employees against predetermined standards, the approach is more reliable than a straight ranking. Nonetheless, since a large amount of comparisons has to be made by the supervisor, it is susceptible for unintentional error. Also it fails to give information about the absolute performance distance between employees and therefore is not suitable for developmental feedback (Shields et al., 2015).

Forced Distribution

A variant of the simple ranking approach is the forced distribution. Applying this approach, the supervisor apportions employees according to a normal distribution function (Aguinis, 2009). The concept goes back to Jack Welch, who during his time as CEO of General Electric introduced the so called 20-70-10 rule. Thus, supervisors had to assign their subordinates to one of three categories with a predetermined distribution. 20 per cent of employees should be rated “superior” (reward and promote), 70 per cent of employees should be assigned to the category “average” (challenge and support) and the remaining 10 per cent as “poor” (dismiss) (Sebald and Jacob, 2015). With the aim of fostering a “true meritocracy”, Welch decimated its workforce each year. In doing so, he aimed at raising the bar of performance continuously (Shields et al., 2015). Whilst this approach is not very popular in Germany (partly due to the fact that dismissal due to one bad performance rating is prevented by labour law), the approach, known – sometimes modified to forced rankings -, is most common in the United States(Armstrong, 2018; Sebald and Jacob, 2015). Some organizations in the United States modified the practice by annually terminating the employment of the five to ten per cent of the lowest performers. Therefore, this practice is often referred to as “rank and hank” (Armstrong and Taylor, 2014).

The approach also is often praised by its supporters to enable a “true” differentiation among employees, allowing management to identify and promote top-performers (Sebald and Jacob, 2015). By this, the approach addresses some of the weak points of simple rankings or rating scales. Further, by forcing supervisors to clearly differentiate between their subordinates, it serves to minimise central tendency or leniency error. There is also some evidence in research (J. Berger et al., 2013) that forced distribution can lead to a rise in performance and productivity, at least in the initial years of application (Sebald and Jacob, 2015; Shields et al., 2015).

Despite its widespread use, especially in the United States, the practice of forced distribution has some major shortcomings. One of the most essential points of criticism is that it is often disliked and perceived as unfair by the employees under review. This is majorly due the perception that the practice is arbitrary and the rating therefore is artificial (Chartered Institute of Personnel and Development, 2016a). Therefore, employees interpret the practice of forced distribution usually as a lack of interest and a lack of appreciation for their work (Chartered Institute of Personnel and Development, 2016a). Moreover, forced distribution supports a harsh climate of competition among employees and teams rather than fostering a climate of trust and high involvement (Shields et al., 2015). Maybe due to those issues, a trend to ditch forced distribution can be observed

in many organizations during the last decade (Chartered Institute of Personnel and Development, 2016a). Even General Electric, who pioneered in applying forced rankings, abandoned the practice in 2005 after Jack Welch left the organization (Cappelli and Tavis, 2016).

c) output-oriented approaches

Target setting

As a results oriented approach, target setting is a refinement of the "management-by-objectives (MBO) concept that was proposed by Peter Drucker in the 1940s (Shields et al., 2015). More current approaches of target setting for evaluating performance were jointly developed by Locke and Latham in the 1990s (Chartered Institute of Personnel and Development, 2016a). According to Latham (2004), there are four causal mechanisms of targets that affect performance: directive function (guiding employee´s efforts towards target-relevant activities), energising (ambitious targets lead to greater effort), increasing persistence (spending prolonged effort to reach a target) and the identification and use of task-relevant knowledge (Chartered Institute of Personnel and Development, 2016a). The rationale behind this approach is that performance will improve if employees are directed towards future achievement rather than being reviewed on past performance.

The target-setting process includes several steps. First, key result areas need to be defined. Subsequently, key performance indicators can be derived for each performance area, informing definitions of performance targets and standards. Finally, an agreement on a set of targets should be reached between the employee and the organization (Armstrong, 2018). For many years, there was broad agreement that participative target setting or even self-assignment would work better to improve performance since employees would be more self-committed. However, currents studies and meta-analysis show that there is strong empirical evidence that assigned targets are more potent since they are tied to some form of external expectations, control or evaluation (Chartered Institute of Personnel and Development, 2016a).

One of the major benefits of target setting is that the approach is rather simple and open-ended in nature and therefore readily adaptable. For administrative purposes it can easily be linked to the payment or bonuses. By applying "objective" performance indicators it is more likely to overcome issues concerning the subjectivity and susceptibility to unreliability inherent in behavioural assessment. Further, the approach

enables an alignment with the strategic objectives of the organisation (Shields et al., 2015).

At the same time, Latham and Locke themselves indicated some potentially serious shortcomings of target setting. One of the most obvious is the temptation to ignore performance-relevant tasks that are not directly assigned to target achievement. Further, there may be target conflicts between employees, which may become counterproductive to organizational performance (Latham and Locke, 2006). Although supporters of the target-setting approach point out that it is more objective than input-oriented approaches, it has to be acknowledged that the definition of key performance areas and indicators as well as the weighting of targets is still subject to human judgement and therefore is prone to subjectivity and error. More recent criticism focuses on the short-termism of (particularly) financial targets and the impact on ethical behaviour, which also opens the possibility employee stress and "burnout" (Shields et al., 2015). Thus, some studies showed that target setting in HPWS often lead to an increase in unethical behaviour (Welsh and Ordóñez, 2014). Therefore, Shields et al. (2015) stress the importance of ethical awareness when using target setting for performance evaluation in highly competitive or high performance environments.

Balanced Scorecard

One approach that supports the strategic alignment of targets is the BSC. Since this approach often is applied on organizational level, it was already discussed in chapter 3.2.5.2.

Visual Performance Assessment

The Visual Performance Assessment is an alternative approach to rating or target setting using visual method of assessment. The basis of the approach is an agreement between the supervisor and the employee deciding on where the latter should be placed on a matrix or grid. The agreement will be recorded in a review document. The aim of this practice is to get a balanced assessment of the employees´ target attainment over an agreed period. The assessment not only can take target attainment into account. Since the matrix consists of two axes, it may also be reflected what competencies were developed (Armstrong, 2015).

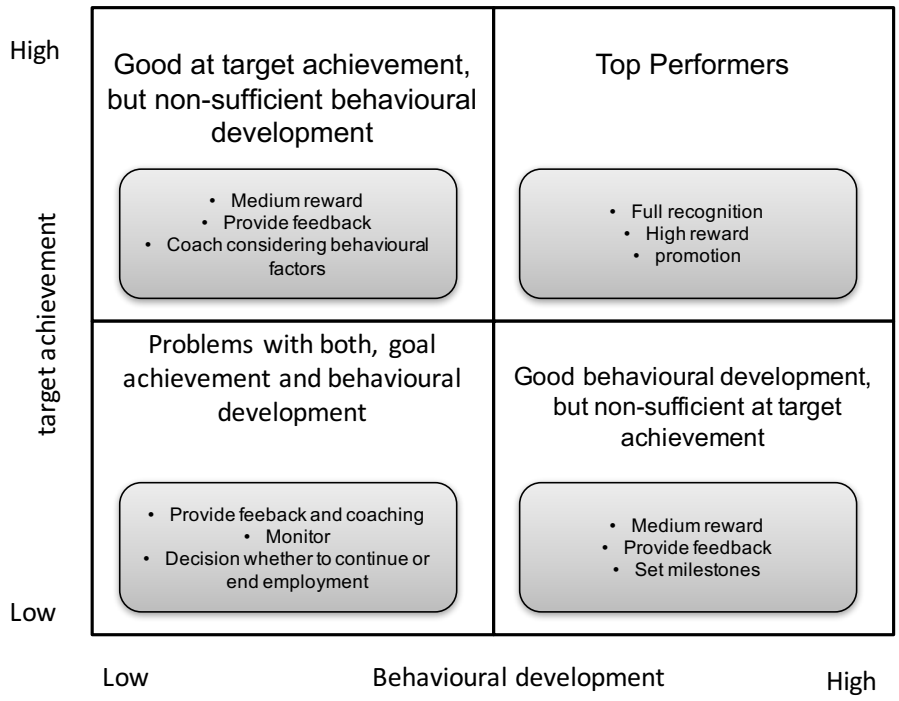


Figure 14: Visual Assessment (modified from Armstrong and Taylor, 2014)

The assessment matrix provides an overview of the employee´s overall contribution, which is presented visually. It therefore can provide a better basis for analysis and discussion than a mechanistic rating. The matrix also provides guidelines on what possible actions for improvement may include (Armstrong, 2015).

One of the advantages of a visual assessment is its simplicity. It is easy to understand and moreover, it provides a sound basis for discussing and deriving developmental needs. Thus it also is more of a forward-oriented approach, supporting employees to identify areas of improvement rather than focusing on short-term results of the past. In this respect, it addresses one of the major shortcomings of the traditional target-setting approach (Armstrong and Taylor, 2014). Also, by including behavioural development on one axis, the visual assessment is able to combine input-oriented and output-oriented approaches.

However, to come to a classification, assessment results for both perspectives need to be provided. Thus, the visual assessment is an approach that hardly stands on its own, but works in conjunction with other approaches like ratings or rankings including their respective shortcomings. Further, if the matrix only displays the individual´s performance results, it holds no information about benchmarks or performance distances to others.

This, however, can be solved by including the performance information of other employees (Armstrong, 2015).

3.2.6.3 Summary of review

As it becomes evident from the review, there are a wide range of PM approaches on individual level, each of which has its own particular strengths and weaknesses. The table below summarizes the key findings for each approach considering its strengths and shortcomings, as well as its reliability regarding intentional and unintentional rating errors.

Absolute Approaches	**Strengths**	**Weaknesses**	**Reliability**	
			In countering unintentional errors	In countering intentional errors
Critical incidents appraisal	- Directly relates to individual behaviour - Allows for individual feedback	- No qualitative measures - Time consuming - May be perceived as intimidating	Moderate - high	Low-moderate
Graphic rating scale	- Simple - Provides absolute measure of performance - Enables identification of low and high performers	- Highly subjective - Trend to oversimplification - Often fail to differentiate performance sufficiently	Low-moderate	Low-moderate
Behavioural Anchored Rating scale (BARS)	- Provides detailed descriptors for behaviour and grading schemes - Focuses on relevant job behaviour - Informs both administrative and developmental purposes	- Other behavioural aspects may be overlooked - Assumes that employees can fully control all performance indicators	Moderate	Moderate
360-degree feedback	- Reduces individual bias - Supports high involvement culture	- Time consuming procedure - Prone to multi-rater bias - Participants often lack accountability for results - Often leaves employees and assessors without guidance	Moderate	Moderate

Comparative Approaches				
Straight ranking	-Simple - Differentiates between top and low performers	- Provides no developmental information - No information about relative distance to others - Highly subjective - Assumes all employees are comparable to one another - Fails to differentiate average performance	Low	High
Paired comparison	- More systematic than rank order - Oriented towards predetermined performance standards	- No information about absolute performance distance - No developmental information	Low	High
Forced distribution	- Differentiates between top, average and low performers - Can lead to increase in performance and productivity in the initial years	- Often disliked by employees (perceived as unfair) - Fosters harsh, competitive climate - Benefits are often short lived	Low	High

Output-oriented approaches				
Target Setting	- Simple - Informs both administrative and developmental purposes - Applies quantitative, objective performance indicators - Supports alignment of individual targets and strategy	- Ignores relevant tasks that are not covered by targets - Fosters target conflicts and stress - May have negative impact in behaviour (morally, ethically)	High	Moderate
BSC	-Focuses on performance indicators that have the highest impact on performance - Includes stakeholder perspectives - Supports alignment of individual targets and strategy	- Unable to assess overall performance (no aggregation of dimensions) - Provides no benchmark information - Assumes linear causal relationships between performance dimensions	High	Moderate
Visual Assessment	- Simple - Provides basis for performance discussion - Future-oriented	- Only works in conjunction with other approaches - Provides no benchmark information	Moderate	Moderate

Table 4: Summary of review of PM Approaches on individual level

Apparently, there is no approach that enables an assessment free from intentional or unintentional rating error, determines relative and absolute performance differences in a consistently reliable manner and does so in a way that is simple to understand, exercise and that is cost-efficient. Whilst all behavioural approaches are highly subjective, absolute approaches are moderately more suitable to prevent unintentional rating errors. Further, they rate better at providing developmental information. At the same time, they are often more complex (except for the graphic rating scale) and therefore more time-consuming and costly. Comparative approaches, in contrast, generally are more cost efficient and allow for a more precise differentiation in performance. They also manage well to control for intentional rating errors.

However, comparative approaches rate poorly in controlling unintentional error and provide only little feedback information (Shields et al., 2015). Output-oriented

approaches, on the other hand, are generally less subjective and therefore more suitable to control for assessor´s rating errors. Moreover, they support the alignment of individual performance and an organization´s strategy. Since they do not reflect on input factors, they might induce unwanted or unethical behaviour to achieve the determined performance targets. Further, most approaches assume that target achievement can be fully controlled by the employees themselves. Since no approach manages to account for all requirements, many organizations base the selection of their evaluation approach on a trade-off and compromise. Others use input- and output-oriented approaches or a combination of both (Shields et al., 2015).

3.2.7 Data Envelopment Analysis for assessing individual performance

The aim of this thesis is to investigate whether the Data Envelopment Analysis, which so far has been used to evaluate the performance of organizations or units, is a suitable approach to evaluate employees´ performance. The rationale and underpinning theory of DEA has already been illustrated in detail in chapter 2. After reviewing traditional approaches for evaluating employees´ performance, this section aims to identify areas where DEA may address some of the shortcomings of the traditional approaches and where some drawbacks of DEA may be located.

As already outlined, DEA evaluates the performance of all units within a given set by converting multiple input and output measures into a single performance score by using linear programming technique. Thus, DEA combines the input- and output-oriented view. Being of comparative nature, DEA provides a detailed differentiation of performance, identifying top-, average- and low-performers. In contrast to traditional comparative approaches, it provides rich feedback information and indicates the relative distance to peers. Therefore, it should be able to inform both, developmental and administrative decisions. Since the performance score is calculated automatically and weights are assigned by linear optimization, DEA eliminates unintentional rating error, which is a common flaw of comparative approaches. As a result of the evaluation, DEA provides a ranking of all employees under review. By comparing only employees that have a similar input-output structure, and therefore are comparable to one another, DEA addresses another shortcoming of traditional approaches. Thus, several employees can achieve the first ranking position. Like absolute approaches, DEA provides a score for overall performance. However, since this score is the result of a complex calculation derived from several input and output criteria, an allegation of “oversimplification” may not hold. Further, for a more differentiated evaluation, a multi-stage DEA, regarding several performance dimensions, could be applied.

Like output-oriented approaches, DEA is based on objective performance indicators rather than on subjective judgements. Unlike these approaches, DEA does account for uncontrollable factors (environmental factors, for instance) acknowledging that there are factors that affect employees´ performance, which they cannot fully control.

A drawback that DEA shares with some of the traditional approaches is its rather complex nature. This makes the evaluation procedure time-consuming and the results may be difficult to interpret without further skills or guidance. In addition, DEA is quite sensitive to outliers and requires - depending on the number of performance measures - a larger set of comparable DMUs or a sophisticated sensitivity analysis.

3.2.8 The impact of HRM and Performance Management on performance

Since HRM in general and performance management approaches in particular, aim at enhancing individual and organizational performance, there has been a lot of research effort during the last decades to investigate whether those approaches perform well and whether they have an impact on performance at all (Armstrong, 2018). Furthermore there seems to be little understanding about the underlying mechanisms through which HR practices influence organizational performance (Wall and Wood, 2005). Hence, this section will discuss evidence on the impact of performance management on organizational and individual performance.

3.2.8.1 Impact on organizational performance

To establish a link between HRM practices - including performance management - and organizational outcome, is problematic since determining causality is a major issue in this field of research(Armstrong, 2018). Nonetheless, there are several research projects and studies that dealt with this question. Generally speaking, there are two main strands of research:

(1) Studies that doubt or even deny any linkage between HR practices and organizational outcome

(2) Studies that support the linkage between HR practices and organizational outcome

The table below provides an overview on studies investigating these two main strands.

Author(s)	Object(s) of investigation	Major findings
(1) Studies in doubt of linkage		
Godard (2001)	longitudinal study of 78 Canadian farms	Found that employment practices believed to complement alternative work practices were not positively related to the perceived gains from this work practices, with some of them even negatively related.
Cappelli and Neumark (2001)	longitudinal data from National Employer Surveys in the U.S.	High Performance practices raise labour costs but do not affect efficiency.
Guest et al. (2003)	performance data of 366 British corporations	There is little or no association between HR practices and performance outcomes.
Wall and Wood (2005)	assessment of 25 key empirical studies published from 1994 onwards including only studies that covered multiple HR practices and that were using measures of economic performance	There are hints that HRM actually does promote organizational performance, the empirical evidence is not strong enough yet to justify that statement.
Studies supporting linkage		
Arthur (1994)	steel mini-mills	There is a proven link between organizations´ strategy of HRM practices
MacDuffie (1995)	explored HR practices in auto assembly plants	There are bundles of HR practices that are related to organizational outcomes such as productivity or quality.
Delaney and Huselid (1996)	the HR efforts of 590 for- and non-profit firms	A positive associations between HR practices, such as training and stuffing effectively, and perceptual firm performance measures was found.
Becker and Gerhart (1996)	case studies on High Performance Work Practices	Empirical work has progressed enough to suggest that HR practices are crucial for "creating and sustaining organizational performance and competitive advantage".
Paauwe and Richardson (1997)	30 empirical studies	There is a relationship between a bundle of HR practices and important HRM outcomes on employee level, such as satisfaction, motivation or commitment. Additionally there is also a link between those outcomes and more general outcomes on organizational level like quality, customer satisfaction or profit.
Boselie; Dietz et al. (2005)	104 empirical research articles dealing with the linkage between HR practices and organizational performance published between 1994 and 2003.	It is impossible to directly compare the results of the different studies, but those studies show that HRM in its "system" has an impact (in a positive sense) for organizational performance.
Messersmith et al. (2011)	Welsh public-sector employees	There is a significant linkage between HR practices and departmental performance. Furthermore, HRM-systems have an important impact on employee attitudinal variables like job satisfaction or commitment.

Table 5: Studies on the linkage of HR practices and performance

Although literature provides no definite answer to the question whether there is a linkage between HR practices and organizational outcome, it becomes obvious that some progress has been made in the last two decades. In spite of the different levels of confidence about the strength of the association, recent reviews reveal that there is a

growing body of research that consistently demonstrated a relationship between and HR practices performance (Paauwe et al., 2013). In this respect, it seems reasonable to suggest that HR practices and organizational outcome are at least weakly associated in a positive way.

Even by assuming a linkage, there is no clear understanding about the mediating mechanisms between HRM and organizational performance. In 1996, Becker pointed out the lack of knowledge about *"the process (how and why) through which HRM creates organisational value"*. Guest (1997) addressed this challenge by emphasizing that literature lacks *"a theory of HRM, a theory of performance and a theory of how the two are linked"*. Purcell et al. (2003) argued that previous studies demonstrated that there is a positive association, but there is no explanation to the nature of this connection.

Those statements, pointing out the existing gap in explaining the link, illustrate an issue, which HRM literature refers to as the "black box" (Boselie et al., 2005). In order to address this issue, several authors attempted to open the box by providing models to map the relationships including intermediary ones, in the HRM-Performance chain (Savaneviciene and Stankeviciute, 2010). Most of these models, which are also known as "causal pathways", are based on the concept of Dyer and Reeves (1995), who defined four different categories of output (Boselie et al., 2005; Savaneviciene and Stankeviciute, 2010):

- HR-related outputs (affective, cognitive, behaviour)
- Organizational outputs (quality, productivity)
- Financial outcomes (profit, sales)
- Market based outcomes (market value)

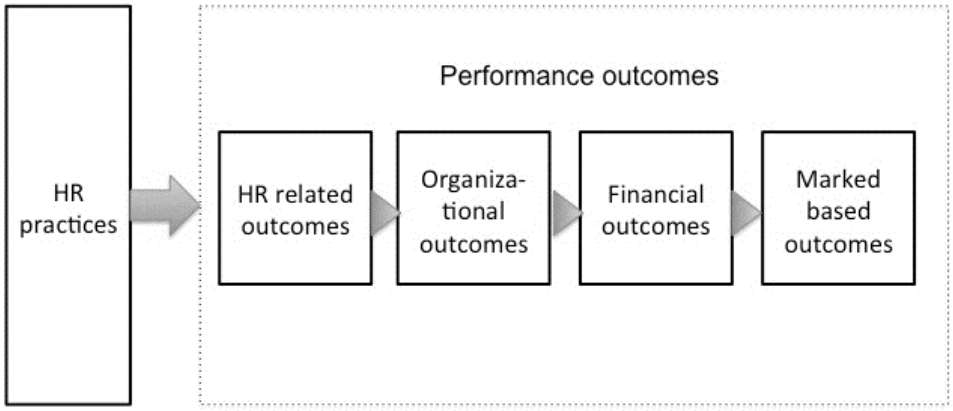


Figure 15: HRM impact on Performance logic (Savaneviciene and Stankeviciute, 2010)

The relevance of the “causal pathway” models lies in two aspects. First, there are outcomes, such as HR-related outcomes, that are closer related to HR practices than others. Second, the impact that HR practices have on more distal outcomes is achieved by the impact on more proximal outcomes (Savaneviciene and Stankeviciute, 2010). Some of the most relevant and frequently cited “causal pathway” models are:

- The Black box model of Becker et al. (1997): HR practices have a direct impact on employee´s skills and motivation, which in consequence influence more distal outcomes.
- The Black box model of Guest et al. (1997): High performance on individual level does not only depend on high motivation, but also on the possession of the necessary skills, abilities and an appropriate role.
- Black box model of Wright and Nishii (2007): Each organization selects HR practices resulting from a HR strategy, which the organization´s management believes will most effectively elicit the employee responses desired. However, not all intended practices are actually going to be implemented exactly the way they were meant. (Purcell et al., 2003).

By comparing the three models, three findings become apparent. First, HR practices result from HR-strategy, which again is derived from the organization´s strategy. Differences in the composition and focus of HRM therefore may result in differences in employee´s attitudes and behaviour. Thus, control oriented HR systems may result in rather compliance oriented behaviour (Messersmith et al., 2011). Second, HRM and HR practices should focus on group or on individual level, as each employee may in fact experience a different HRM system. More recent studies provide evidence that there is a relationship between employee well-being and organizational performance. This line of research indicates, that understanding how HR practices impact individuals is a prerequisite for understanding the relationship of HRM and performance (Paauwe et al., 2013). Third, HR practices impact HR-related outcomes. Again, those outcomes influence more distal outcomes such as financial performance. Although the “causal pathway” models do not completely unlock the “black box”, they offer a reasonable basis to argue that to have a positive impact on performance and HR-related outcomes do have an impact on organizational performance.

3.2.8.2 Impact on individual performance

In contrast to performance management´s impact on organizational performance, there is convincing evidence of a causal link between performance management and a positive impact performance improvement on individual level (Armstrong, 2018).

Of course, there are also studies that found no such linkage. Armstrong (2018) reports on a study carried out by Guest and Conway (1998) who used achievement of financial targets and skill development as key criteria for determining the impact of performance management approaches. Although they found that ninety per cent of respondents rated performance management as “moderate” to “highly” effective, a more detailed statistical analysis produced no convincing evidence of any link between performance management practice and target achievement or skill development.

However, there is a large number of studies that provide convincing evidence that there is a link. In a comprehensive review of more than 3,000 research articles, Kluger and De Nisi (1996) showed that feedback had a moderate-sized positive impact on performance. They, however, pointed out that results had a strong variation which might be due to different approaches. Further, Kochanski (2007) found that high-performing organizations have a strong leadership support for performance management. Despite identifying several issues with performance management practices (see chapter 3.2.9), in a comprehensive review the CIPD identified several studies indicating a positive relationship of performance management and individual performance (Chartered Institute of Personnel and Development, 2016a; Chartered Institute of Personnel and Development, 2016b). Regarding the practice of target setting, Corgnet at. al (2015) found that employees who found their targets challenging increased their performance by forty per cent compared to the control group. Jeffrey et al. (2012) showed that ability-based targets were more suitable to improve performance results than a “one-size-fits-all” approach. From a reversed point of view, Menefee and Murphy (2004) noted that strong performers are more attracted by organizations that recognize individual contributions.

In conclusion, there is evidence to believe that there are benefits of performance management to individual performance. However, there are also critical voices on the practice and the context performance management operates. These issues will be addressed in the following section.

3.2.9 Issues with Performance Management

Although several studies have proved that PM can help to improve organizational and individual performance, many PM related initiatives failed to deliver on their promises (Micheli and Mari, 2014) . In consequence, there is criticism in HRM and PM research on whether PM is beneficial to organizations and their employees. Hence, the criticism, is not only addressed at the way performance is evaluated, but it doubts the benefit of evaluating employee performance in general.

Two aspects that are commonly cited when it comes to general criticism of performance management are the so called “Hawthorne effect” and the “Red Beads parable”. The Hawthorne effect was first observed in a study initiated in 1924 by the management of the Hawthorne plant of the Western Electric (in cooperation with the Massachusetts Institute of Technology (MIT) and Harvard University). They wanted to investigate the relationship between illumination, rest pauses and work hours on productivity. In a series of runs they found that performance increased steadily, but the investigated variables did not explain the increase. Thus, the researchers concluded that the improved personal relationship between workers and management that developed during the studies was the real reason for the increase in productivity. The conclusion of behavioural change due to the awareness of being observed soon gave birth to the term “Hawthorne effect”. Since then, the term has been widely used and has mutated in meaning over time and across disciplines.[6] In the context of performance management, results of the Hawthorne studies are often cited to argue that having behaviours assessed engenders beliefs about the assessor´s expectations. Thus, not the results of a performance review but conformity and social desirability considerations lead to change in performance (McCambridge et al., 2014). Since the study was first published, it was controversially discussed. For instance, there have been several subsequent studies trying to reproduce the effect with varying results. Further, many researchers claim methodical flaws (e.g. omitting other uncontrolled variables) and ideological bias in the research (McCambridge et al., 2014). Taking all this into account, it can be concluded that there may not be a “Hawthorne effect” in performance evaluations per se. However, assessors should be aware that there might be specific variables that affect the outcome that are not subject to evaluation.

The parable of the Red Beads was introduced by W. Edward Deming to demonstrate how results are influenced more by the system than by the individual employee. The participant´s task was to “produce” white beads by dipping the paddle into the container

[6] A comprehensive review on the term´s use is provided by McCambrige, Witton et al. (2014) and Chiesa, Hobbs (2008)

of 4,000 wooden beads (of which 800 are red and 3,200 are white) to extract 50 beads at a time. Deming established a factor error rate of only two beads per paddle. In the course of the experiment, the participants took four turns each. Predictably, no one was meeting the quota and despite all advice, reviews, encouragement and criticism by "management" (other participants) the error quota did not improve. Each turn, a different participant was high performer and low performer. Moreover, although each of the used participants the same paddle, the variation in results was different for each worker. However, the results of the experiment indicated a daily average of red bead defects could be calculated. Based on this calculation reasonable upper and lower performance levels could be established, recognizing that failures (red beads) are made and will be made. Deming´s point was to illustrate that in the red beads case, workers had no control over their production no matter what management advised them. The obvious solution was to better manage the material supply in order to avoid to many red beads coming in, which is the responsibility of management not the production workers (Burke, 1991). From the Red Beads Parable, Deming and other researchers draw several conclusions. First, variation is present in any process or operation and the knowledge about one source of the system variation. Hence, it is the system rather than their individual skills that determines their performance. Further, there are always employees who perform above and below average. However, their position in the ranking may vary from one period to the other (Burke, 1991). In summary, Deming employed the Red Beads parable to support his theory about process versus people management, prompting management to turn their focus away from managing employee performance to rather managing interrelated processes of the whole (The Human Resources Social Network, 2001).

Other than criticism on the general benefit of PM, there is also a controversial discussion on its theoretical foundations and its execution in the organizational context. Micheli and Mari (2014) argue that PM still suffers from a lack of underpinning theories and a more rigorous investigation of PM issues could inform current debates. They further claim that, since the measurement of properties in PM is very complex, there is a tendency to treat those indicators as important that happen to be accessible for measurement. Instead, it should be acknowledged by all PM stakeholders that performance is often complex and therefore difficult to define and measure. In consequence, complete empiricity and objectivity should not be considered a necessary condition but rather a goal and the related presence of errors and uncertainties should be admitted and properly dealt with (Micheli and Mari, 2014).

Another issue that has particularly been raised during the last years is PM´s role during the GFC. Even before the GFC, several authors claimed that PM is largely focused on results as opposed to behaviours and competencies (Gruman and Saks, 2011). With emphasizing the output-oriented view, many organizations oriented their PM towards "shareholder value" reducing the holistic approach of PM to a rather narrow focus on short-term financial goals (Pohl, 2015). This shift not only led to work intensification and morally questionable behaviour (see chapter 3.2.6.2) but also to a disconnection from any sustainable HR practices (P. Thompson, 2011).

Regarding PM on individual level, an issue that is constantly brought up by critics is that many employees and supervisors find the procedure too bureaucratic and often not relevant for their jobs. As reported by Adler et al. (2016) the Corporate Leadership Council found that managers spent about 210 hours (employees spent 40 hours) respectively on PM activities. Armstrong (2018) lists a series of studies conducted between 2002 and 2015 which in summary show a rather negative perception of performance evaluation among employees. Among the most cited issues were there lack of perceived fairness, the lack of managers´ skills to execute performance evaluation and the consistency and quality of the approach. Adler et al. (2016) point out that disagreement among assessors and conflicting purposes of performance evaluation are also common issues related to PM. Due to the growing dissatisfaction with traditional approaches, several organizations started to change their performance management system during the last five years, some of them even abandoned it at all (Cappelli and Tavis, 2016).

Despite all criticism, which mainly is accepted among HRM and PM researches, there is also a broad consensus that performance evaluation can be an important and valuable task. Acknowledging all shortcomings, Adler et al. (2016) argue that performance evaluations still have many merits for improving organizations, that "too hard" should not be an excuse to abandon it at all and that the alternative (no performance evaluation) may even be worse. In their comprehensive analysis on the benefits of individual performance evaluation the CIPD (2016a) concluded that there is strong evidence that it is a worthwhile process. However, both studies emphasize, that performance evaluation is not beneficial per se but can do more harm than good if designed and executed poorly. Thus, it is crucial to understand under which specific conditions PM practices can actually deliver the promised results and how the whole procedure should be conducted properly. In this respect, research has identified a number of requirement and practical issues for a proper performance evaluation, which will be discussed in detail in the following chapter.

4 Designing a Performance Evaluation System

A Performance Evaluation System (PES) is a rather complex system that combines the structural and procedural elements of the performance evaluation. Furthermore, it can serve different purposes including administrative, strategic or developmental purposes. To serve its intended purposes, the performance standards and the measures they are assessed with need to be designed properly and underpinned with the required performance information. When introducing a PES, contextual considerations should be addressed too. In the following chapter, the characteristics of and the requirements to a PES are illustrated. Further, the design of an PES regarding the distinct steps of the "performance cycle" are illustrated and contextual considerations are discussed.

4.1 Purpose and stakeholders

Before discussing the design of a PES it should be acknowledged that PES´ may serve different purposes since, in general, there are several stakeholders to the evaluation of employee performance, who do not necessarily share the same expectations or intensions. Hence, the different stakeholders and varying purposes of a PES will be investigated in more detail in the following sections.

4.1.1 The Stakeholders of a PES

With his widely recognized publication "Strategic Management: A Stakeholder Approach" Freeman (1984) established the beginning of what today is understood as "stakeholder theory". The general rationale of the stakeholder theory is the recognition of the fact that several stakeholders are voluntarily or involuntarily contributing to the performance and the success of the organization (Post et al., 2002). Although this concept supplements and enhances the resource-based view and gained wide recognition among researchers and practitioners in business-related areas like accounting, finance and marketing, it has not received much attention from the field of HRM and PM yet (Freeman et al., 2010).

More recently, in the evolving field of sustainable HRM (see chapter 3.1.2), research suggests that more consideration should be put on stakeholders. Thus, several researchers emphasized the importance of considering stakeholders' impact and wellbeing while still achieving financial outcomes for the organization. Mariappanadar (2014) developed a stakeholder harm index, based on a framework for capturing and assessing the externalities of HRM practices on an organization´s stakeholders. Kramar (2014) provides a comprehensive model for sustainable HRM, pointing out that HR

practices cause effects on four levels: on organizational level, on individual level, on social level and on ecological level. To evaluate organizational performance, adequate measures should address the impact of stakeholders from different levels. Thus, sustainable HRM challenges the supremacy of one stakeholder´s interest. For evaluating individual performance, the measures need to be cascaded down to all employees (Kramar, 2014). To identify all relevant stakeholders, several authors provided overviews including Ferrary (2009), Cohen (2012) and Kramar (2014). While the overviews differ considering classification levels or the number of stakeholders, it becomes evident, that they can generally be grouped into internal and external stakeholders. Since the affiliation to the category of external stakeholders depends on factors like the economic environment, the industry and the level of competition, the provided overviews are of limited usability for the individual organization. Merely, they provide an orientation.

However, there is broad agreement, that management, supervisors and employees are generally among the internal stakeholders. In organizations where employees interests are represented by a workers´ council, their members also form a group of internal stakeholders (E. Cohen et al., 2012). Performance Evaluation Systems traditionally were focused on the interests of managers and supervisors. In consequence, research into PM and evaluation of different PM approaches was mostly limited to managers´ perceptions whilst employees´ interests and attitudes often were neglected. However, more contemporary approaches should adopt a broader view, taking into account a variety of stakeholders´ interests among them those of their employees. Therefore, research also needs to focus on incorporating the views and experiences of the wider workforce (Keeble-Ramsay and Armitage, 2015; Kramar, 2014).

4.1.2 Purposes of a PES

An organization may pursue different aims by implementing a PES. On a strategic level, the organization aims to achieve its business objectives by linking organizational performance to organizational objectives (Aguinis, 2009). On individual level, organizations typically use the PES information for two main purposes: administrative and developmental purposes (Kondrasuk, 2012; Meyer, 1991; S. L. Thomas and Bretz, 1994).

Administrative purposes focus on the use of a PES to inform administrative decisions such as pay decisions, decisions on job reassignment, promotion or rewards. Performance evaluations also serves developmental purposes including the provision of performance feedback as a basis for a joint analysis of strengths and weaknesses, for identifying areas of improvement and for individual development plans or learning

contracts (Armstrong and Taylor, 2014) . The figure below provides an overview on the major purposes and their inherent sub-purposes.

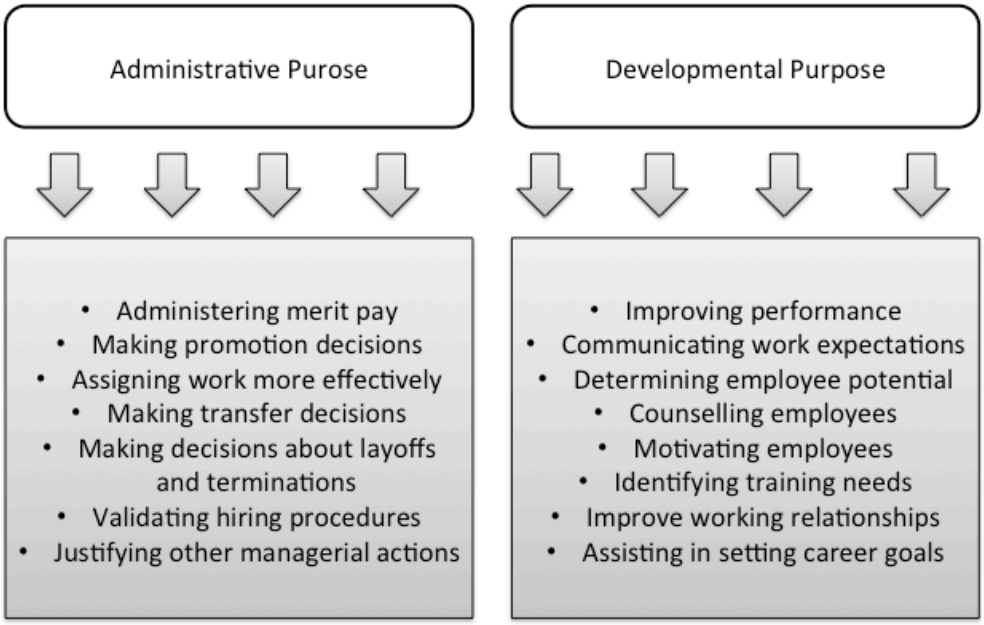


Figure 16: Purposes and uses of PES information (Kondrasuk, 2012; Meyer, 1991; S. L. Thomas and Bretz, 1994)

As noted by Shields (2015), the relationship between developmental and administrative purposes is frequently troubled. Among the most discussed issues is the conflicting role of supervisors if they employ PM information for both purposes. Acting on the administrative purpose puts the supervisor in the role of a judge in a parent-child-type of exchange. By using the performance evaluation for developmental purposes, the supervisor adopts the role of a counselor. Trying to fulfil both roles simultaneously will create conflicts in the relationship of supervisor and employee and therefore is incompatible. A person that is being judged is very likely to hide potentially damaging information and will act rather defensively. On the other hand, employees tend to openly admit weaknesses that could be rectified when the supervisor clearly acts as a counsellor (Kondrasuk, 2012). In a study observing nearly 100 appraisal interviews, Meyer (1991) reports that whenever salary decisions and suggestions for performance improvement were communicated during the same interview, employees´ defensive reactions were very common. They actually were so powerful that attempts to counsel the employee after the salary decision was communicated were mostly futile (Meyer, 1991).

A more recent issue is the change in priorities during the last decade that has been noted by several HR scholars. As Cappelli and Tavis (2016) illustrate in a historic narrative, the

priorities of performance evaluation have regularly shifted over the past decades which prompted different PM approaches. They argue that the current calls for abandoning traditional performance evaluation entirely (see chapter 3.2.6) are a result of changing priorities considering the performance evaluation´s purpose. They claim that advanced economies have a greater need for development and agility rather than for individual accountability (Chartered Institute of Personnel and Development, 2016a). There is indeed a lot of current research indicating that developmental purpose should be privileged over the administrative purpose (Shields et al., 2015). Considering the administrative purpose of performance evaluation, Star et al. (2016) report on a study by Eker and Eker (2009) who found that organizations which employ PM exclusively for monitoring and legitimization usually have a control-oriented top-down culture. Further, many authors claim that using individual performance evaluation as basis for determining performance-related pay or rewards can be the cause of serious practical and ethical concerns – as it was the case in precipitating the GFC (Cascio and Cappelli, 2009; Shields et al., 2015). In contrast, Takeuchi et al. (2009) list a number of studies that found that developmental focused performance evaluations are indicative of motivating employees and foster the perception of a supportive work environment. Further, Kim and Holzer (2016) argue that the developmental use of performance evaluation contributes to trusting relationships between employees and supervisors and heightens the acceptance. Hence, a developmental-oriented performance evaluation seems to be the preferred approach today. This assumption was confirmed by a 2009 UK survey of performance management practices by the CIPD asking “what other HR processes ought to link to performance management?”. 85 per cent of the respondents opted for “learning and development”. In 2004, only 71 per cent of respondents opted for this category. Similar results were found in a study with Australian participants (89 per cent of respondents indicating the determination of training and development as an important purpose of their performance management practices) (Shields et al., 2015).

However, despite the rise of relative importance of the developmental purpose, in their comprehensive study the CIPD (2016a) found no evidence that would suggest to drop performance evaluation for administrative purposes entirely. Rather, they conclude that both purposes are valid. Performance discussions serving different purposes should be separated and it should be clear to employees when each purpose is occurring. For developmental purposes, like potential improvements or discussing immediate concerns, more frequent check-ins are helpful. For administrative purposes including discussions of past performance and how this will affect pay or promotion, annual (or less regular) meetings should be scheduled (Chartered Institute of Personnel and Development, 2017).

4.2 Requirements to a PES

In order to develop a method to properly assess employee´s performance and to design an effective PES, it is crucial to define what features constitute an adequate PES. During the last decades, considerable research has been carried out on what constitutes a good performance evaluation and therefore literature holds a variety of recommendations on this topic. In this context, several authors already provided comprehensive overviews of the most popular recommendations (Armstrong, 2018; Folan and Browne, 2005) The recommendations listed below present a summary drawn from the publications of Armstrong (2018), Cocca and Alberti (2010), Bretz andThomas (1994), Lusch and Serpkenci (1990), Pettijohn and Pettijohn (2001), Folan and Brown (2005) and Aguinis (2009).

According to their findings, an appropriate Performance Evaluation System should (be):

- Perceived as fair
- Free from bias
- Provide clear, accurate feedback
- Linked to compensation/ used for determining rewards
- Use balanced criteria (accounting for uncontrollable factors/ uses input and output factors)
- Open to discussion
- Stimulate continuous improvement/right behaviour
- Linked to the organization´s Business strategy

As already stated, this list is a summary rather that a complete collection of the most agreed requirements to a PES. Whilst the importance to link the PES to the organization´s strategy has already been discussed previously, the other recommendations are addressed in the following sections.

4.2.1 Importance of perceived fairness

In order to engage themselves and feel comfortable, employees need to trust their organization and supervisors and need to feel treated fairly and justly (Gruman and Saks, 2011). The majority of studies investigating evidence on the benefits of PM highlight the fact that performance evaluation needs to be accepted and perceived as fair by the

employees to improve performance or increase job satisfaction (Cocca and Alberti, 2010; Greenberg, 1986; Meyer, 1991; Sudin, 2011). Otherwise, performance is actually likely to decrease (Chartered Institute of Personnel and Development, 2017). Moreover, employees may even deny the accuracy of the whole PES if they do not perceive trust and fairness in the process (Kondrasuk, 2012). When asked about their three major concerns regarding the assessment of employee performance, U.S. organizations identified only issues related to fairness. The three most named issues were (S. L. Thomas and Bretz, 1994):

a) The acceptance of the evaluations system by those being assessed
b) Whether the employees perceive the process as fair
c) Whether the employees believe the results are fair

In this context, several studies confirm that organizations who are attaching considerable importance to the fairness of the performance evaluation are on the right track. The studies indicate that employees´ perceptions of fairness strongly affect their attitude towards job satisfaction, turnover intentions and even workplace behaviour (Sudin, 2011). Gupta and Kumar (2012) showed that there are positive associations between employee engagement and their perception of the performance evaluation´s fairness.

Pettijohn and Pettijohn (2001) conducted a study among 115 salespeople in order to explore the relationship between features of the evaluation process and the resulting level of the salesperson job satisfaction. The results showed that employees experience the greatest levels of job satisfaction if they understand the criteria used and if they believe that the evaluation is fair. Additionally, the impact of the evaluation´s results on compensation had a significant statistical effect on the salesperson´s job satisfaction (C. E. Pettijohn et al., 2001). Another study presented by Mani (2002) confirms those findings. Mani conducted a survey among employees of East Carolina University to review their currently applied PES. The regression analysis of results indicated that fair treatment is a crucial element to the overall satisfaction with the PES. Furthermore, employees´ job satisfaction was significantly related to their perception that the evaluation process was fair and there is no apparent bias in rating. As reported by the CIPD (2016b), in a before-after study, Jawahar (2010) found perceived fairness to positively affect employee´s reaction to feedback and their overall job performance. This finding was confirmed by Budworth et al. (2015) demonstrating that employees´ perception of fairness affects the relationship of feedback and task accomplishment.

But what makes a performance evaluation a fair one? Perceived organizational justice in HRM, particularly in a PES, consists of three distinctive types of subjective

perceptions. These are generally referred to as distributive justice, procedural justice and interactional justice (Sudin, 2011). Distributive justice refers to the relative ratio of an employee´s input to its perceived outcome (Greenberg, 1986). The second type of justice suggests that the fairness of the appraisal process itself is important, apart from the ratings received and therefore is referred to as procedural justice. Thus, procedural justice focuses on the employees´ attention to the procedure of making decisions with the requirement that this process is a fair one. According to Tang and Sarsfield-Baldwin (1996) procedural justice is influenced by issues such as two-way communication, trust in the supervisor, clarity of expectations and understanding of the performance appraisal process. Thus, employees may refuse to agree to evaluation results (or decisions based on those results) because they have no understanding of the process. Further, Kim and Holzer (2016) demonstrated that employees´ perception of performance evaluation in terms of both procedural and distributive justice is significantly and positively related to the developmental use of performance evaluation.

Although there is a strong agreement among researchers that all types of justice have a major impact on the perceived fairness of the PES and result in a higher level of job satisfaction, it is not agreed, which type has the larger impact. Exploring employees´ satisfaction with the PES in Malaysian companies, Sudin (2011) found that particularly distributive and informational justice were significantly related to perceived fairness with the PES. On the other hand several authors stress the importance of procedural justice for meeting evaluatees´ expectations of fairness (Kim and Holzer, 2016; Kondrasuk, 2012) Therefore, it seems appropriate to consider all types of justice when implementing a PES.

4.2.2 Influence of rater´s bias

To base the performance evaluation process on a non-biased assessment not only is a prerequisite for fairness (Kline and Sulsky, 2009; Kondrasuk, 2012) but it also holds the advantage of bringing managers into a counseling mode, rather than bringing them to serve as a judge. However, during the performance evaluation process managers and supervisors have to make a number of judgements. To begin with, they have to define performance standards and performance objectives. In a next step they have to gather information on which the performance is assed. Next, they have to compare expected performance to actual performance and finally have to decide on rewards, promotions or need for training. Those judgements may be biased for a number of reasons (Gentry et al., 1991). Due to favouritism and subjectivity supervisors may give either satisfactory or unsatisfactory appraisals to employees who do not deserve them. In this respect,

supervisors tend to be more apt to give a superior evaluation to someone they actually like (Kondrasuk, 2012). In other cases, biased judgements result from shortcomings in people´s ability to process information and therefore are classified as “unintentional errors”. Research findings about unintentional errors arise from decision theory. They are mainly based on the research of Kahnemann and Tversky (1977) , who referred to unintentional errors as “judgemental heuristics”.

4.2.3 Provision of feedback information

Both, the developmental and the administrative purpose of a performance evaluation presume another important requirement: the performance evaluation must be able to provide clear feedback information and direction, which enables counselling on how to improve performance (K. Becker et al., 2011; Latham et al., 1993) as well as decision making about promotion or rewards. Therefore, obtaining feedback information is one of the most important requirements to a performance evaluation (Armstrong and Taylor, 2014). However, if feedback is not reasonably accurate, for instance inconsistent or unreliable, it may not be accepted by the employees and their subsequent performance may even deteriorate (S. Adler et al., 2016). In their comprehensive review on evidence of performance evaluation, the CIPD (2016a) also confirms that feedback generally contributes to improve performance, but that there is a great deal of variety. Thus, if executed poorly it has no effect or worsens performance (Chartered Institute of Personnel and Development, 2016a).

Although there is no “universal template” to follow in order to obtain accurate feedback information, there some courses of action that were found to be helpful by several researchers. Perceived fairness and reduction of bias are found to be among the most important issues (Chartered Institute of Personnel and Development, 2016a). To be accepted and acted upon, feedback information should not only be easy to understand but also it should enable an equitable treatment of different groups of employees. The information needs to be specific enough to precisely point out need for improvement as well as to give positive reinforcement to those employees who excel (Mani, 2002). A major issue in practice with feedback information is that supervisors, once the evaluation is completed, are overcharged with how to use the information effectively. Due to a lack of standardization, supervisors are uncertain about how to implement the results (Kondrasuk, 2012).

Another important issue to feedback information is that there is strong evidence that it is rather the reaction to feedback than the feedback itself that determines how feedback affects performance (Chartered Institute of Personnel and Development, 2016a). For

instance, Murphy and Cleveland (1995) showed that psychological and behavioural reactions establish the extent to which employees use the given information to alter their performance. Further, Smither et al. (2005) found that employees who express rather positive emotions after they received feedback continue to obtain higher performance results that employees who experienced negative emotions. Aguinis et al. (2012) argue that feedback should focus on employees´ strengths rather than on weaknesses to enhance employee engagement. Those findings are a strong case for applying rather strength-based approaches to obtain feedback, for consulting employees on a regular basis on issues such as performance evaluation and for being mindful about their response and perception of the evaluation (Chartered Institute of Personnel and Development, 2016a).

4.2.4 Linkage of performance results to pay and to target achievement

If the PES is designed to adequately reflect different levels of performance, the results can also be employed to determine levels of compensation. Although often requested by managers and employees, this claim is controversially discussed in literature and practice. Thus, Smither (1998) emphasizes the necessity of a performance-based pay-plan since employees expect rewards to be linked to the evaluation process. Ilgen and Feldman (1983) argue that evaluation results could even negatively affect job satisfaction, when the linkage between performance results and compensation is not made explicit. Another argument that supports the linkage of the performance evaluation to compensation is that all participants treat the evaluation process with more seriousness if rewards are a part of it. In their study among 115 salespeople Pettiijohn and Pettijohn (2001) showed that linking performance evaluation results to compensation had a significant statistical effect on the salesperson´s job satisfaction. Correspondingly, they conclude that performance evaluation results should be linked to pay and that managers should discuss the reward implications to increase employee´s perceptions of a pay-performance relationship.

A case-study conducted by Mani (2002) draws other conclusions. Written comments that were added on the survey by participants emphasize that pay is important to the perception of the system´s effectiveness, but that the system employed caused dissatisfaction due to skewed distributions. Mani (2002) quotes:

> *“The System rewards all employees rated “better than good” the same. Employees are not motivated to do any better than good to get raises.”*

"It is not the "method" of evaluation that discourages many, it is the poor reward system."

Those statements reflect the general challenge concerning the linkage of performance evaluation to compensation. It is not that employees generally refuse to be paid or rewarded by a performance-based method. It is rather the method itself that is subject to discussion and often the cause of dissatisfaction. Thus, whenever performance evaluation results are used as a basis to determine compensation, the underlying measurement method should be able to identify good and poor performers in an objective (non-biased) and fair manner. In addition, managers need to communicate compensation implications in advance and subsequently act on them once the performance evaluation is completed.

The use of target setting as an output-oriented PM approach has already been discussed in chapter 3.2.4. To improve performance, the linkage of performance evaluation to target achievement is crucial (Chartered Institute of Personnel and Development, 2016a). As stated by the CIPD (2016a), Harkin et al. (2016) found that monitoring the progress towards target achievement is important for motivating employees towards specific attainment. Moreover, Neubert (1998) showed that adding feedback information to target setting contributes to performance improvement and target achievement (Chartered Institute of Personnel and Development, 2016a).

4.2.5 Use of balanced criteria

To design a systematic approach to performance evaluation on an individual level, it needs to be defined in advance what skills, knowledge and ability are required and what outcomes are expected. According to Shields (2015), a PES should measure one or more of these variables: inputs (competencies), the way these inputs are applied (behaviours) and the outputs (results) that are obtained. Other authors reduce those evaluation bases to two major categories – input and output bases (Jackson Jr. et al., 2010; L. S. Pettijohn et al., 2001). A comprehensive analyses of input and output criteria frequently used in sales force performance evaluations is provided by Jackson et al. (2010).

Studies on the usage of evaluation criteria do not indicate a consensus what criteria should be preferred. Some researchers argue that output criteria should be preferred over input criteria suggesting that output criteria are more objective. In addition, some argue that outputs such as sales and profit are "the only things that count". On the other hand, output based methods are unsuitable to provide adequate feedback information

about training needs or career improvement. To account for both, administrative and developmental purposes, a performance evaluation subsequently should incorporate inputs as well as outputs. This would also enable conclusions about efficiency, which – due to efforts in maximizing productivity – had become an important component of performance evaluations recently (Cook and Zhu, 2006; Jackson Jr. et al., 2010).

An adequate evaluation of performance also needs to consider that every employee may face a slightly different set of environmental characteristics (Grifell Tatjé and Marques-Gou, 2008; Kondrasuk, 2012) that may affect their performance, but which they have no control of (Gentry et al., 1991). This emphasizes the role of inner- and other –direction (or locus of control). The locus of control describes the degree to which individuals perceive events in their lives as a consequence of their own actions and decisions (inner-direction, controllable) or as not being related to their actions and decisions (other-direction, beyond their control) (Lusch, 1990). A PES that is able to incorporate both input factors that are controllable and input factors that are beyond the employees´ control may not only identify areas that the employee needs to improve, but also areas where the organizations needs to make changes relative to their inputs for employees (Cook and Zhu, 2006). Furthermore, the inclusion of non-controllable factors surely may increase perceived fairness.

4.2.6 Stimulation of continuous improvement/right behaviour

An issue, which often is overlooked, is that performance measures always have a behavioural impact. Systems that involve humans respond to performance measures. Thus, employees are likely to modify their behaviour in an attempt to ensure a positive performance outcome. If performance measures are not designed properly, this may lead to actions that are inappropriate and contrary to the company´s strategy. The measures themselves are less a problem than the disregard of the behaviours they may induce (Neely et al., 1997). The reduction of customer complaints, for instance, is a reasonable objective for a service organization and the decrease in complaint rates may be a subsequent measure derived from this objective. If employees´ performance is evaluated applying the measure “rate of complaints” it is very likely that they try hard to reduce their number of complaints, including not asking the customers for feedback. Thus, the organization would miss crucial hints for improvement and may even dissatisfy customers who try to complain. Hayes and Abernathy (2007) take this argument one step further by arguing that inappropriate performance measures and poorly designed incentive schemes were to blame for a short-term U.S. business culture in the late 1970s. Many HR-academics argue, that poorly defined performance measures and short-term

performance targets may also have contributed to the recent GFC (Keeble-Ramsay and Armitage, 2015; Psychogois et al., 2016; P. Thompson, 2011)

Therefore, the design of performance measures should not be a one-time event, but a continuous process to ensure, that the assigned measures stimulate the right behaviour and that they are still appropriate to reflect the company´s objectives (Neely et al., 1997).

4.3 The Performance Management Cycle

Regarding the design of a performance evaluation system (PES) there are generally two aspects to take into consideration: the content of the evaluation (what is evaluated) and the process of the evaluation (how is it evaluated). The content of the evaluation again depends to a large extent on the purposes of the PES (Fletcher, 2001). Considering the process of performance evaluation there is a variety of different models (an overview is provided by Gruman and Saks (2011)). In general, these models consist of a sequence of stages corresponding to Deming´s P-D-C-A circle (Armstrong, 2010; Gruman and Saks, 2011). The process starts with planning activities based on the agreement on purposes, performance standards and performance targets. To foster employee engagement, these indicators should be subject to negotiation (Gruman and Saks, 2011). The stage is completed when agreement is reached on how performance will be measured, what is expected and what evidence will be used to establish level of performance (Armstrong, 2010). The agreement stage is followed by performance execution and monitoring activities with a focus on job design, coaching and supervision. The next step focuses on the assessment of performance usually based on the results of a specific performance management approach (as discussed in chapter 3.2.4). The results analysis should bring an agreement on the level of performance achieved and provide a basis for development and improvement in the subsequent performance review. Also, to promote engagement, the assessment procedure should focus on perceptions of justice and trust (Gruman and Saks, 2011). During the performance review stage, results should be discussed with the employee. For developmental purposes it should not only be considered if performance targets have been achieved but also where individual strengths and weaknesses and areas for improvement lie (Armstrong, 2010). When pulled together, individual performance results should also inform decisions on corporate level. Therefore, both performance evaluation on individual and corporate level need to be aligned (see chapter 3.2.4).

Although the basic steps of the evaluation process are fairly common, a PES requires to take a holistic perspective comprehensively addressing the constituents of performance

(Gruman and Saks, 2011). Thus, Folan and Brown (2005) suggest to combine the structural and procedural elements into a holistic framework. Whilst the structural elements are concerned with the configuration and make-up of the evaluation itself (including performance standards and measures or the frequency of the evaluation), the procedural elements reflect the distinct steps of the performance evaluation cycle.

By enhancing the procedural approach developed before, the figure below illustrates how the structural elements (in the white boxes) can be integrated into a procedural framework.

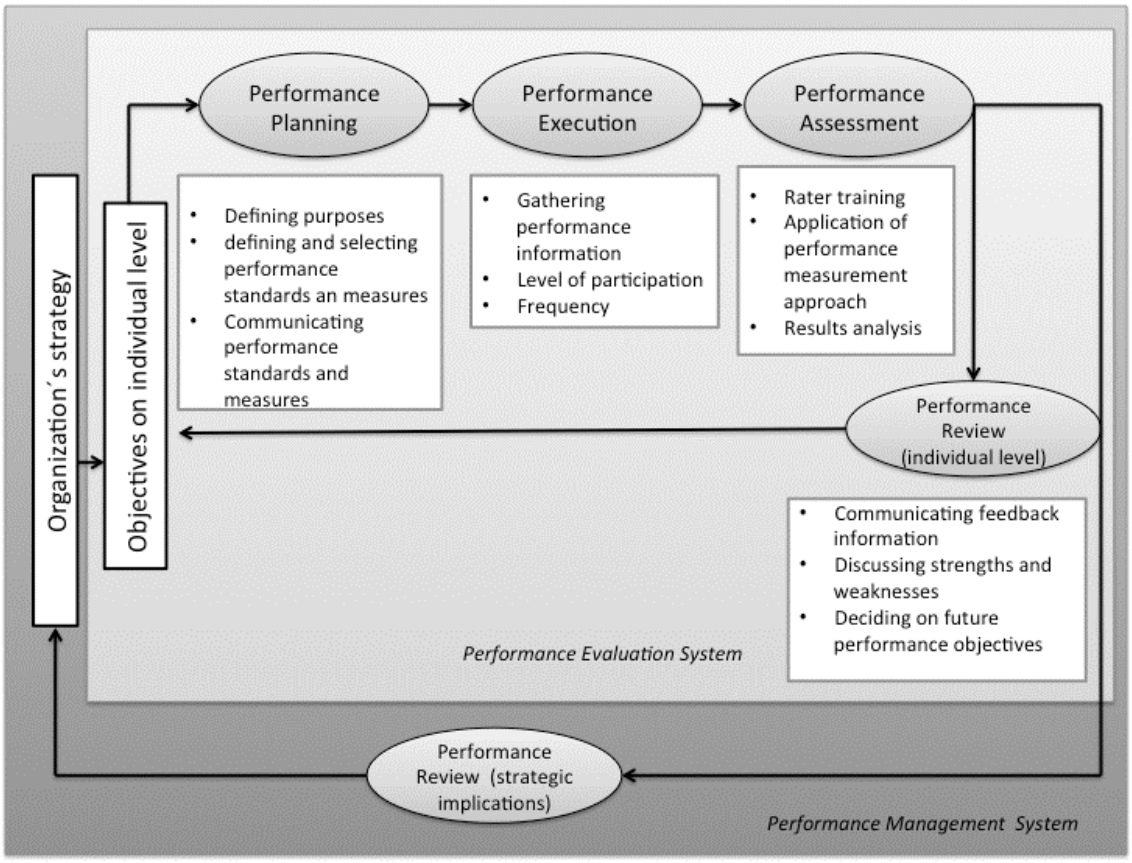


Figure 17: Structural and Procedural elements the PM cycle

In the following sections, the performance management process including its structural elements will be examined in more detail.

4.3.1 Performance Planning

One first important task in the design of a PES is to agree on the purposes and targets of the evaluation. An essential step in this process is to establish targets that, on the one hand, reflect organizational objectives but on the other hand also address values and

personal objectives of employees. This way, a PES is more likely to be integrated with the employees´ selves and to result in experiences of meaningfulness and employee engagement (Albrecht et al., 2015). Criteria or rules about the number of objectives are hardly provided. For reason of economy and utility, most authors suggest to concentrate on only few objectives (Star et al., 2016).

The next two steps, namely the identification of performance dimensions (often referred to as key performance indicators) and the definition of clear and understandable performance measures should proceed in parallel. These steps are essential to align the performance measures with performance targets (Star et al., 2016). Further, the influence of bias becomes less, whenever clear standards and observable performance information is used. Another important issue is that employees should understand what is being measured and how it is done (Fletcher, 2001). To enhance adaptability, most of the cited frameworks are open and can be interpret in various ways. Thus, any single organization can decide which measures to capture under each of the headings (Neely et al., 2000). This leads to the question, what defines and adequate measure and how can executives decide on which performance measures they should rely on?

In order to identify key features, which determine a “good” or “adequate” performance measure, Neely et al. (1997) analysed a variety of different papers and books on performance measurement. The literature was not randomly chosen, but was based on a more extensive review in one of Neely´s earlier studies (Neely et al., 1995). As a result, Neely et al. (1997) provided a comprehensive summary of recommendations with regard to the design of performance measures. The five most suggested recommendations are:

No. 1: Performance measures should be derived from strategy/ match the organizational context.

No. 2: Performance measures should be simple to understand.

No. 3: Performance measures should provide timely and accurate feedback.

No. 4: Performance measures should be based on quantities that can be influenced or controlled by the employee alone or in cooperation with others

No. 5: Performance measures should reflect the business process.

In an attempt to encapsulate those recommendations, which together constitute an “adequate” performance measure, Neely et al. constructed a framework for specifying performance measures. The framework is referred to as “performance measure record sheet” and consists of ten elements (Neely et al., 1997).

The figure below shows how these elements are combined into a performance measure record sheet using the example of the measure “new business won”:

Title	New Business won
Purpose	Business Objectives – „achieve sales targets“, „increase market share“ and „achieve financial returns“
Target	100 per cent of contracts targeted
Formula	Orders received versus order targeted (expressed as percentage)
Frequency	monthly
Who measures?	Sales manager
Source of data	Order book
Who acts on the data?	Sales director
What do they do?	Identify shortfalls in performance and investigate reasons
Notes and comments	This measure requires to introduce a new management process – namely the screening of contract opportunities

Figure 18: performance measure record sheet (Neely et al., 1997)

Although the design of measures is a process, the performance measure record sheet is more of a structural element, providing guidance to support this process. Thor (1993) developed a more operational guideline and proposes nine general rules for designing performance measures:

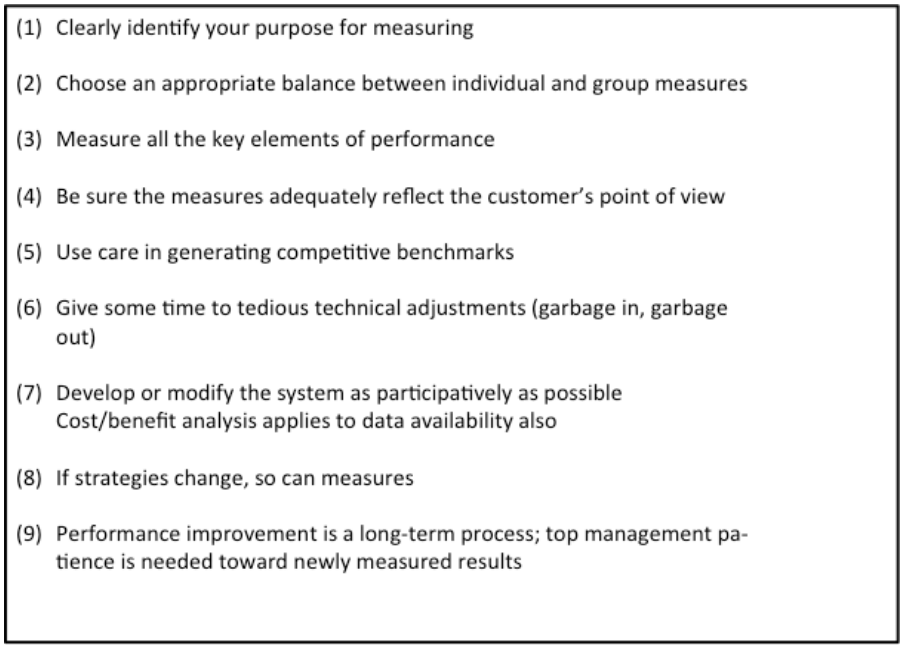


Figure 19: Guidelines for designing performance measure (Thor, 1993), adapted from (Tangen, 2004)

In summary, the approaches of Neely et al. (1997; 2000; 2000) and Thor (2003) should be considered valuable guidelines for the design of performance measures.

Considering the quality of measures, Micheli and Mari (2014) argue that measurement in social science is a process that aims at "adequate-to-purpose" results rather that at "true" results. Hence, by replacing the term "true" with "adequate", decisions over cost and quality become an essential decision in the measurement process. The authors therefore suggest to follow a pragmatic paradigm by acknowledging that measurement results are informational entities and that the trade-off between acceptable quality and availability of data or resources is the most general criterion for "good measurement".

4.3.2 Execution of Performance Evaluation

The integrity of the performance evaluation – regarding not only the measurement approach, but also what information is used and how it is gathered and processed – is crucial for acceptance and perceived fairness of the whole PES. Thus, performance information should be based on explicitly defined sources of data (Globerson, 1985). Those sources may include supervisors, peers, subordinates, self or customers as well as documents such as sales reports. Thus, having better knowledge about what information is gathered and who contributes information, the more employee acceptance is likely to increase (Aguinis, 2009). Accordingly, to reduce liking and bias, clear and observable performance information should be used, whenever available (Varma et al., 1996). Kondrasuk (2012) additionally suggests the inclusion of a measure that reflects

the situation, e.g. the economic situation or the specific characteristic of an assigned territory. Moreover, Star et al. (2016) suggest to check the collected data for accuracy, validity and completeness on a regular basis to ensure that all stakeholders of the process can be confident in the credibility of the information.

Another constant line of research regarding the execution of the evaluation process deals with the participation level. In this respect, Cawley et al. (1998) suggest a distinction between instrumental participation (for the purpose of influencing an end result) and value-expressive participation (for the purpose of having one´s voice heard). They found the latter to have a stronger relationship to employees´ response to the evaluation than the first. In two different surveys among salespeople, Pettijohn et al. (2001) found that participation in the evaluation process is an integral part to employee´s job satisfaction and organizational commitment. Roberts (2003) also identified the participation level to be essential to any fair and ethical PES (Levy and Williams, 2004).

An adequate timing and arrangement of the elements of the evaluation process are also an important and widely discussed issue. Traditionally, most organizations conducted an annually evaluation by setting performance objectives at the beginning of each year. Then, the employee is given time to perform and to meet those objectives. At the end of the year, the performance is assessed, a meeting to discuss the results takes place and the cycle continues. Considering the different purposes of the evaluation process and the supervisor´s associated role conflict of being a judge and a counsellor at the same time, it is recommended to separate review meetings (Chartered Institute of Personnel and Development, 2016a). Whilst in traditional evaluations those meetings were conducted annually, there has been a major change taking place towards executing developmental meetings more frequently and less formally (Cappelli and Tavis, 2016).

To disseminate the results of the assessment to the employees and other stakeholders, a user-friendly, easy comprehensible reporting format should be selected. The reporting format is essential for demonstrating results, building awareness and supporting accountability. Further, it may help to identify unforeseen issues and to enhance utility of the indicator data. Regardless of the specific format, limiting the number of indicators included and the adherence to graphic design principles should be key to a user-friendly format (Star et al., 2016).

4.3.3 Performance Assessment

As described before, the evaluation of employee performance involves many stakeholders. Therefore, the successful application of the chosen performance

management approach and the execution of the evaluation process require a clear understanding of how the system works. Furthermore, the benefits of including the different perspectives of all involved stakeholders should be emphasized. Thus, a communication plan addressing the following issues should be implemented (Aguinis, 2009):

- What is performance management and how does is fit into our strategy?
- How does it work?
- What are the benefits?
- What are the responsibilities of each person involved?

Before disseminate the results of the evaluation to the employees and other stakeholders, the results should be analysed in detail considering aspects like the meanings of the findings or whether performance moves in the desired direction. If necessary, supplemental information needs to be gathered to fully understand the results. Therefore, the analysis of results should take place prior to the review meetings with employees. If dissemination fails, subsequent actions based on the results of the PES may be questioned (Star et al., 2016). To help preventing dissemination failure and to provide assessors with tools that allow them to conduct the evaluation process effectively, both researchers and practitioners emphasize the importance to engage in rater training. Trainings should address issues including to be able to explain the PM approach and the underlying measures, how to conduct an interview or how to coach employees helping them in planning their progress (Aguinis, 2009; Kondrasuk, 2012).

4.3.4 Performance Review

The cornerstone of each evaluation process is the performance review, which generally is executed by interviews between supervisor and employees (Aguinis et al., 2012). As already noted, to foster employee engagement and motivation, the evaluation needs to be perceived as fair and just. If employees lack trust, performance reviews are unlikely to produce positive results. As noted by Kahn (1990), situations that promote trust are “predictable, consistent, clear and nonthreatening”. Thus, an essential requirement to any performance review meeting is that there is clarity on the purpose of the interview (administrative or developmental, for instance) on all sides. Further, supervisors should provide feedback on their performance based on evidence (Armstrong and Taylor, 2014; Gruman and Saks, 2011).

After implementing a PES, an ongoing monitoring and evaluation of the system should not be neglected. Since this is the only way to respond to negative perceptions of stakeholders, to bad quality of results or to changes in the organization´s business strategy, the monitoring process should be conducted on a continuous basis. Evaluation data should include reactions to the system as well as an assessment of the systems operational and technical requirements (Aguinis, 2009). To assess reactions to the PES, a confidential survey among the employees during the initial stages of implementation could be administered, addressing perception of fairness and acceptance towards the system. This could be repeated after the end of the first cycles to find out if there have been any changes. To assess management´s point of view, interviews with managers and supervisors who have been involved in developing and implementing the system can be conducted. To assess the quality of the system´s results, performance ratings should be assessed over time to analyse if the systems has positive impacts (Harper and Vilkinas, 2005).

4.3.5 Considerations to the design of a PES

The structural and procedural elements of a PES need to be integrated effectively in order to link individual performance to organizational performance. To affect and improve individual performance and foster employee engagement several recommendations that research found to be positively related to have been discussed in the previous sections. However, the successful implementation of a PES is also majorly dependent on the contextual environment within these recommendations are employed (Star et al., 2016).

As already illustrated in chapter 3.1.3 and 3.2.9, there are many critical voices on HRM and PM criticizing these academic fields for removing from their original concern for the well-being of individuals and even for having contributed to the recent global financial crisis (GFC). Thus, Thompson (2011) claims there is no realistic way to reform HRM as long as liberal market economies are dominated by shareholder value logic. Thus, he argues, greater methodological sophistication would be welcome but not sufficient to rescue a “failed project”. According to Thompson (2011), a change may only be brought by external regulation of employment systems. Less definite, but just as pragmatic, Star et al. (2016) point out that power relationships between stakeholders and conflicting stakeholder interest are two major issues that may lead even a well-designed PES to failure. Both issues may result in a power-imbalance. By analysing several case studies, Bitici et al. (2006) found that organizational culture and management style can shape a PES. They provided examples where organizational cultures evolved from a dominant

top-down to a rather consultative culture and therefore concluded that management styles need to evolve as culture evolves (Star et al., 2016).

In conclusion, it should be noted that the success and the achievement of the intended aims of a PES not only depend on proper design but also on management style, the power relationships between stakeholders and other contextual environmental factors. Thus, if the organization employs a style of control and command or privileges shareholder interests over employees´ interests, the PES is very unlikely to support employee engagement or to improve performance even if designed properly. On the other hand, if designed and executed poorly, a performance evaluation may also fail in open, engagement and learning oriented cultures. Thus, organizational context and design need to be consistent.

4.4 Contextual considerations for the case study

As pointed previously, the benefit and success of a PES not only depend on a proper design but also on the context of the application. Therefore, some essential contextual issues, in particular on the labour market and political framework conditions as well as the legal context of the case study are discussed in the following sections.

4.4.1 Labour market and political framework conditions

While most studies on HRM assume that their findings apply universally, a growing body of research shows that there are differences in the way HRM is conducted in different countries (Brewster, 2007). Thus, there is a lack of evidence that there is a common global HRM archetype. Rather, several studies identified a range of framework conditions and regulations of the labour market that lead to different ways in which the impact of HRM on institutions and employees can be understood (Wood et al., 2012). Since this thesis illustrates a case study that was carried out in a German institution, the differences in the way HRM is understood and conducted in Germany and the UK shall be examined more closely.

In research, there are different approaches to explain the variations in the way HRM is conducted. Among the most commonly cited is the “relationship” approach within the varieties of capitalism (VoC). This approach emphasizes the difference of formal political frameworks and institutional features such as labour market institutions or the severity of competition policy (Wood et al., 2012). It distinguishes between two models of capitalism: liberal market economies (LMEs, e.g. the USA or the UK) and coordinated market economies (CMEs, e.g. Germany or Japan). LMEs are characterized by a market

dominant logic and a strong focus on shareholder value. Thus, in UK organizations, managers are particular sensitive to shareholder´s demands and face high pressure to prioritise short-term shareholder values over employee interests. Also, workers´ councils are rather marginalized (Kang and Moon, 2011). Although this is more so in the US than the UK, the unions play a weaker role in the UK, for instance negotiating the rules governing internal labour markets, compared to unions in Germany. Moreover, there are few legislative restrictions inhibiting employers´ pursuit of enhanced organizational flexibility (Tregaskis and Brewster, 2006). There is also a strong preference for voluntarism considering issues like corporate social responsibility or sustainability (Kang and Moon, 2011). In consequence, this leads to weaker employee rights and low employment security.

In CME´s, on the other hand, organized interests including business associations and workers´ councils play a dominant role. Thus, organizations in CME´s tend to have a stronger focus on stakeholder-value. Hence, managers are not only sensitive to the demands of shareholders but have to balance the interests of all stakeholders. In conclusion, employee rights are stronger and job security is higher than in LMEs. In Germany, stakeholder participation is supported by a formal and legal structure of co-determination where representatives of workers´ councils sit on the board (Kang and Moon, 2011). Further, through the use of secure employment contracts, strong wage levels and employment protection against changes to working conditions, employees are encouraged improve their skills and stay loyal to companies that invest in them (Tregaskis and Brewster, 2006). The GFC brought to light the limitations of the LME-style shareholder value oriented model challenging the USA and UK as archetypes of this model. Although there have been some regulatory reforms, there has been no attempt yet to warrant a path-shifting change (Kang and Moon, 2011). However, there is a growing consensus that there is a need for a more balanced view on stakeholders´ interests and for greater accountability.

4.4.2 Legal considerations

The design and implementation of a PES may also touch legal issues. This applies in particular, if the results of the performance evaluation are used to determine pay or if they are used for decisions about promotion or dismissal. Thus, a sound performance evaluation process will help to avoid legal liability (Smith, 2013). In general, a PES that is fair and acceptable to employees is likely to be also legally sound. The basic principles that underlie the implementation of a PES should therefore include procedures that are known by everyone involved and that are applied in the same way to everyone (Aguinis,

2009). The table below lists characteristics of a legally sound PES illustrating those principles:

Characteristics	That should (be)
Performance dimensions and measures	Job related, clearly defined and explained to the employees
Procedures	Standardized and uniform to all employees within a job group (depending on the structure of the organization separate review formats might be developed for different job groups)
Assessment and review	Timely and periodic
Objective measurement	Cleary identify which results constitute acceptable performance and which indicate performance below organization standards
Employee involvement	Enabled through possibility to comment on procedure and results, self-assessment
Documentation	Thorough and consistent, including specific examples

Table 6: characteristics of legally sound PES (modified from (Malos, 1998; Smith, 2013))

There are also several laws that affect the implementation and the execution of a PES. Within the last decades, many countries around the world have passed laws prohibiting discrimination based on race, sex, religion, disability status or sexual orientation. In this context it is essential to distinguish legal from illegal discrimination, since adequate PES´ should be able to discriminate among employees based on their level of performance. If a PES could not do this kind of legal discrimination effectively, it would be rather useless. However, an appropriate PES does not discriminate illegally in terms of like age, sex or ethnicity (Aguinis, 2009).

In the United Kingdom the following laws have been passed to prevent illegal discrimination (Aguinis, 2009):

- Equal Pay Act (1970)
- Sex discrimination Act (1975)
- Race Relations Act (1976)
- Disability discrimination Act (1995)
- Employment Equality Regulations (2003)

In Germany, the German General Equal Treatment Act, which aims to prevent or eliminate discrimination on grounds of racial or ethnic origin, gender, religion or belief, disability, age or sexual identity passed in 2006. The scope of this act also refers to recruitment and selection conditions as well as to employment and working conditions (German Federal Ministry of Justice and Consumer Protection, a):

In addition, the German Works Constitution Act of 2001 addresses operational co-determination. Thus, section 82 enacts the employees´ right to be heard on any matter concerning his person. Furthermore, the employee can request an explanation concerning the calculation of pay or ask for an interview on the assessment of his performance. Section 94 (2) enacts the approval of the works council to the formulation of general assessment criteria (German Federal Ministry of Justice and Consumer Protection, b).

Regarding variable remuneration, an ordinance on the supervisory requirements for German Financial Institutions' Remuneration Systems was developed in the aftermath of the 2008 financial crisis. The so called "Remuneration Ordinance for Institutions" came to force in 2013. For the requirements to a remuneration system sections 4 and 5 (1) apply in particular.

> *Section 4: "The remuneration parameters shall be aligned with the strategies and support the achievement of the strategic aims"*
>
> *Section 5: (1) Remuneration systems are appropriately designed if incentives for the members of the management body and staff to take disproportionately high risks are avoided.*

Taken together, these laws urge organizations in Germany to implement performance evaluation systems that treat all employees equally regardless of demographic or ethnic characteristics. Furthermore, all employees have the right to request an explanation of how their remuneration is calculated and how their performance is assessed. Thus, an adequate performance measurement approach should provide clear and accountable information about that. Finally, workers´ council has to agree on the general assessment criteria and therefore should be involved in or at least informed about the process of defining performance dimensions and measures. To calculate remuneration, the remuneration system needs to be properly designed and should be aligned with the organization´s strategy.

5 Managing Banking Services

Banking services are an integral part of the service landscape of each developed society. Although, in research, there is an ongoing discussion about the existence of distinctive features, which distinguish services from goods, it is widely agreed that some services require management approaches that go beyond traditional techniques. This is especially true for banking services.

In the scope of this chapter, the evolution of service management is outlined. In this context it is also discussed whether there are unique characteristics that distinguish services from goods and if they are applicable to all types of services. Subsequently, the role of human labour in the production of services is reflected. Finally, banking services are illustrated in more detail. Since the case study that is illustrated throughout this thesis was applied in a German cooperative bank, this is done with a special focus on the German banking sector. Eventually, banking services are analysed with respect to their inherent service characteristics. Based on this analysis, challenges for managing banking services are derived.

5.1 The Evolution and Definition of Services

Since services touch the lives of each person in developed economies, service sector research is an immensely growing field within academic literature. Although activities of manufacturing and agriculture will always be necessary, service markets have never been larger, competition in services has never been more intense and the economic growth within developed economies is almost exclusively based on services (Fitzsimmons and Fitzsimmons, 2008; C. Lovelock and Gummesson, 2004). Thus, the service sector in the United Kingdom meanwhile accounts for more than 77 per cent of the economy (Inman, 2015). The picture is similar in Germany, where service industries generated about 70 per cent of economic value in 2015 and currently contribute up to 75 per cent to employment (German Federal Ministry for Economic Affairs and Energy, 2015). Recent economic statistics show that, while the share of customers services did not shift significantly over last years, the share of business services lie finance, renting and R&D did. Thus, they are the key drivers for economic growth (Wirtz and Ehret, 2017).

The identification of services as the central focus for economic exchange goes back to a number of early economic philosophers and scientists. More than 150 years ago, Bastitat (1848/1964) claimed that *“the great economic law is this: Services are exchanged for services”*. Walras (1894/ 1954) broke the term *“services of capital goods”* down into services that have direct utility (“consumer services”) and services that have

only indirect utility ("producer services"). He reasoned that the failure to include the immaterial services of capital goods prevented the development of a pure research area (Vargo and Lusch, 2008). As many of the early economists identified the primacy of service in exchange, Delaunay and Gadrey (1992) summed up their work by stating *"(…) they reckoned that one should not study whether activities are productive or not, but how they interrelate"*. Penrose (1959), one of the first authors to support the resource-based view of the organizations, claimed, that *"it is never resources themselves that are "inputs" to the production process, but only the services that the resources can render"*. During the 1970s and 1980s, service researchers put forward the message that services required distinctive management practices. They argued that, until then, these were not addressed by research, which was mainly grounded in manufacturing and agricultural production (collectively referred to as "goods"). After internal debates within academic institutions, a consensus emerged, indicating that services have different characteristics than goods and that these characteristics pose vexing management challenges (C. Lovelock and Gummesson, 2004). In a comprehensive study, reviewing 46 publications by 33 authors, Zeithaml et al. (1985) identified the four most cited distinctive features of service: intangibility, inseparability, heterogeneity and perishability (often referred to as "IHIP" characteristics), which will be illustrated in more detail in the following section. Within this logic, services are intangible outputs of an organization – hence, intangible goods. More recently, researchers argue that, although IHIP still applies to a majority of services, the division between goods and services is outdated. They point out that existing service concepts do not apply to the variety of services existing (Vargo and Lusch, 2008), arguing that it is not only four distinctive features that separate services from goods. Moreover, service is a term to capture the process of using resources to the benefit of some entity. Thus, a service is a multifaceted construct that is the centre of economic and social exchange (C. Lovelock and Gummesson, 2004; Vargo and Lusch, 2008).

The evolution of the service concept becomes also visible considering its different definitions over time.

Service Definition	Source
"A service is an activity (…) of more or less intangible nature, that normally, but not necessarily, takes place in interactions between customer and service employees and/or physical resources (…) which are provided as solutions to customer problems."	(Grönroos, 1990)
"A service is a process or performance rather than a thing."	(C. H. Lovelock, 1991)
"Services are deeds, processes, performances."	(V. Zeithaml and Bitner, 2003)
"Service is defined as the application of specialized competences (operant resources and skills) through deeds, processes and performances for the benefit of another entity or the entity itself."	(Vargo and Lusch, 2008)
"A service is support for an individual's or organisation's everyday processes in a way that facilitates (or contributes to) this individual's or organisation's value creation."	(Grönroos and Gummerus, 2014)

Table 7: service definitions

Although the definitions are on a rather abstract level, it becomes evident, that their scope has broadened. Whilst earlier service definitions emphasized how services are different from goods, more recent definitions reflect the perspective of services. However, the various definitions of "services" are open to interpretation. Thus, it is not clearly defined what is meant by "activities, deeds or processes". In addition, there is a wide range of perspectives, e.g. emphasizing the role and the benefit of the customer or rather a perspective on value creation. Thus, there is no precise definition of service (Edvardsson et al., 2005). A common denominator of most definitions is that services are (or consist of) "activities" or "processes" that are applied or done to the benefit of an entity (Vargo and Lusch, 2004). Thus, the definition according to Vargo and Lusch (2008) will be applied throughout this thesis.

5.2 Distinctive Characteristics of Services

As already stated, the relationship of goods and services caused a wide-ranging debate within service sector research. The discussion particularly focused on how services characteristically differ from goods and what these differences imply for managing services. By conducting comprehensive reviews, several researches tried to identify unique service characteristics. Rathmell (1966) identified thirteen characteristics; Lovelock (1991) identified seven. However, the four most cited characteristics, that are commonly accepted, were identified by Zeithaml et al. (1985) based on a review of 46 publications. Although other characteristics have also been suggested, the IHIP characteristics have found near-uniform and almost unquestioned acceptance by researchers over a long period of time (Vargo and Lusch, 2004). Those characteristics are

- Intangibility: Since services are no physical objects, they cannot be touched, seen, felt or tasted (Edvardsson et al., 2005).
- Heterogeneity (lack of standardization): Since no two customers are precisely alike, their expectations and demands will also be rather unique. Thus, it is very challenging to impose standards for the production of services or for service outputs (C. Lovelock and Gummesson, 2004).
- Inseparability: service provider and customer interact simultaneously to produce a service. Following this logic, a service cannot be produced away from the customer (Vargo and Lusch, 2004).
- Perishability: Since services are not tangible like goods, it is often claimed that they cannot be produced in advance, stored for later usage, returned or resold (V. Zeithaml and Bitner, 2003).

More recently, the validity and representative nature of those characteristics have been questioned. Edvardson et al. (2005) claim, that IHIP characteristics are based on observations and practical experience rather than on empirical research. Vargo and Lusch (2004) note, that by first assuming commonly accepted characteristics of (tangible) goods and subsequently defining service in terms of absence of these characteristics, the IHIP characteristics represent a definition by exclusion. Thus, they are often interpreted as negative qualities or hurdles to overcome. Additionally, there is a growing agreement among scholars, that IHIP characteristics fail to delineate service

from goods adequately in some cases (Edvardsson et al., 2005; C. Lovelock and Gummesson, 2004; Vargo and Lusch, 2004).

5.2.1 Generalizability of service characteristics

The discussion about the universality of the IHIP characteristics reveals that there is at least some serious doubt that those characteristics are able to capture the essence of services and provide a clear guideline to distinguish services from goods. In a study that was carried out by Edvardsson et al. (2005), eleven service scholars were asked if they felt that the generic IHIP characteristics are able to capture the essence of service. As a results, nearly half of the experts (five out of eleven) claimed that IHIP does not portray services in a meaningful way. Six experts found that IHIP still serve as a useful description, for they still capture some of the essence of services, but that they are not as universal as originally posted. Thus, Edvardson et al. (2005) concluded, that IHIP might not be generalizable to all services, but that they should be used to some services when they are relevant and in situations when they are useful.

Gummesson and Lovelock (2004) share this view. They conducted a comprehensive literature review to identify research projects that have investigated IHIP characteristics by studying the complex properties of all types of services. When they found no such study, they concluded that the claim, that services are different from goods based on IHIP, is not supported by empirical evidence. Furthermore, they indicate that IHIP only apply for certain types of service. Based on those findings they segmented services into four subcategories and applied IHIP to each category. The categories are based on whether the service is a physical act in nature and whether humans themselves, owned objects or information are the central element that is processed to create a service.

	Service Category Involving				
Characteristic		**Physical Acts to customer´s body (e.g. healthcare)**	**Physical Acts to owned Objects (e.g. car repair)**	**Nonphysical Acts to customer´s Minds (e.g. education)**	**Processing of information (e.g. insurance, accounting)**
	Intangibility	Misleading – performance is ephemeral, but experience may be tangible	Misleading – performance is ephemeral, but physically transform possession may be tangible	Yes	Yes
	Heterogeneity	Yes – often hard to standardize due to direct labour and customer involvement	Numerous exceptions – can often be standardized	Numerous exceptions – can often be standardized	Numerous exceptions – can often be standardized
	Inseparability	Yes	No – customer usually absent during production	Only when performance is delivered live	Many exceptions – customer often absent during production
	Perishability	Yes	Yes	Numerous exceptions – performance can often be stored electronically or printed	Numerous exceptions – performance can often be stored electronically or printed

Table 8: Applicability of IHIP characteristics to different types of services (C. Lovelock and Gummesson, 2004)

Sixteen cells result from the combination of four service categories and four service characteristics. An evaluation of the cells reveals that there are numerous exceptions from the claim that all services inhibit each IHIP characteristic. Thus, some researchers even reject the goods/ service divide and conclude that both are interdependent (Grönroos and Ravald, 2011; Gummesson, 2014). However, it becomes obvious that IHIP characteristics lack universality when it comes to distinguish services from goods. Nonetheless they are very helpful to define a certain type of service.

5.2.2 Service Classifications

In an attempt to differentiate different types of services and addressing subsequent management issues, several classification schemes have been proposed. Chase (1978) suggested to array services along a continuum from high to low contact, with “contact” referring to the customer´s presence in the service system. Maister and Lovelock (1982) enhanced Chase´s scheme by adding another dimension: the extent of service customization. For Maister and Lovelock (1982) service customization activities involved

compiling a tailored service for each customer. Applying two dimensions to classify services resulted in a two-by-two matrix yielding a four-way classification. The authors labelled the resulting categories “service factory”, “service shop”, “mass services” and “professional services”. Since Maister and Lovelock did not provide a comprehensive description of the four service types, other authors modified the two-by-two matrix by adding other dimensions, which they felt were more suitable to capture the essence of distinct service types. Thus, Schmenner (1986) claimed that services were better classified using both the degree of labour intensity and the degree to which the customer interacts with the service and the degree to which the service is customized for the consumer.

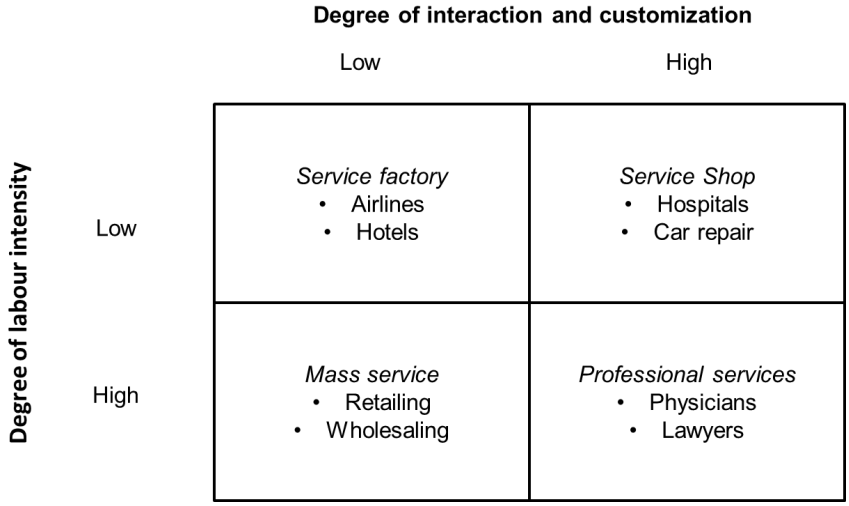


Figure 20: The service matrix (adopted from Schmenner (1986)

The figure above displays Schmenner´s service matrix and indicates some service businesses that fit into one of the four quadrants. The “service factory” is characterized by a low degree of labour intensity and a low degree of interaction and customization. Thus, service providers operate in a fashion similar to factories. They can take advantage of economies of scale, may employ rather unskilled employees or automat their services. With an increasing degree of interaction, the “service factory” gives way to the “service shop”. Service providers operating in this fashion still have a high degree of equipment relative to (human) labour, but offer more interaction and customization. Exhibiting a low degree of customization, but being very labour intensive, “mass services” include the rather “traditional” service businesses like retailing and wholesale. With an increasing level of customization or interaction, mass services give way to “professional services” (Schmenner, 1986).

In order to classify the service process as a whole, Silvestro et al. (1992) somewhat combined the schemes of Chase (1978) and Schmenner (1986). Thus, they proposed a two dimensional approach. One dimension is the volume of customers per unit or the customer contact time. The other dimension consists of six sub-dimensions drawn from the service operations literature: contact time, level of customization/ standardization, discretion, people/ equipment focus, front or back office orientation and process or product orientation. As a result, Silvestro et al. postulate only three different types of services: “mass services”, “service shops” and “professional services”. Unlike Schmenner, they position the distinct service types along a diagonal, reflecting the service process:

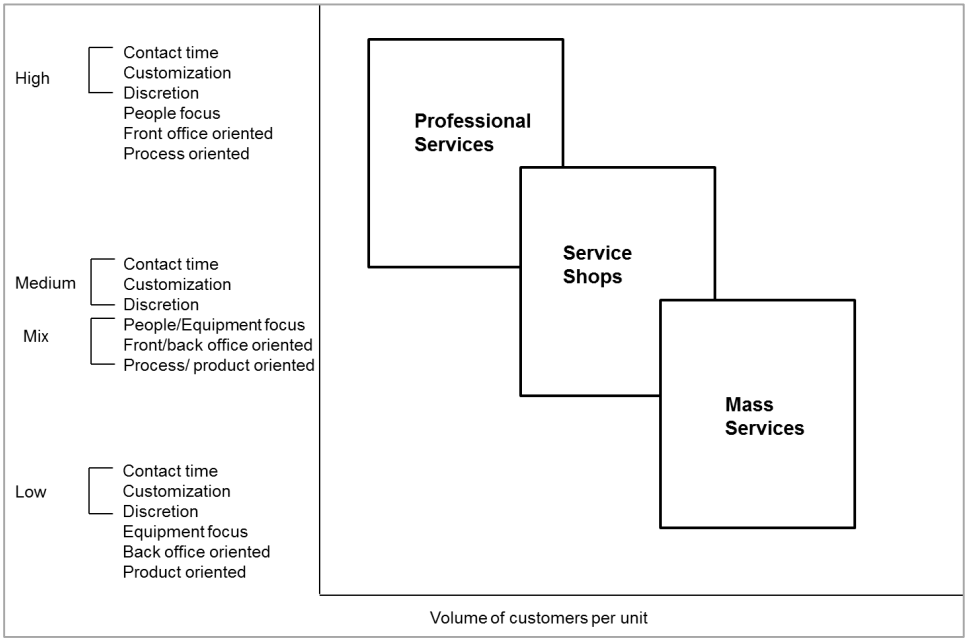


Figure 21: Model of service processes (Silvestro et al., 1992)

Although Silvestro et al. (1992) applied the same terms as Schmenner (1986), the definitions of the distinct service types differ slightly. According to Silvestro et al. (1992), “professional services” are highly customized services with a rather long contact time. Respectively, customer transactions per unit are rather low. Thus, professional services are often highly specialized. “Mass services”, on the other hand, are characterized by little contact time and little customization. Hence, there are many customer transactions. It can be noted that this definition of mass services combines Schmenner´s definition of “service factories” and “mass services”. Eventually, “service shops” fall between mass services and professional services with the levels of classification dimensions falling between the two extremes (Silvestro et al., 1992).

5.3 The Role of Human Labour in the Service Sector

Even if the IHIP characteristics are not applicable to all types of services, the foregoing discussion clearly shows that human labour is an important factor in the production of services. Particularly during the interaction with customers, employees play an important role and can influence the result and the quality of the result significantly (Benkenstein et al., 2017). Although there are also attempts to replace human inputs by automation in the service sector, these endeavours do not always lead to economic success. In an attempt to cut costs and still produce the same amount of services, many service providers try to restructure their resources, for example banks using ATMs or urging their customers to use their mobile phones and PCs to take care of regular bank transactions. If customers perceive the result to have the same or even better quality as before, these changes have been successful and the organizations´ revenue generating capability has clearly improved. However, the replacement of human labour by automation in the production of services can also have the opposite effect and result in service processes with the perceived quality of the output deteriorating (Grönroos and Ojasalo, 2004). In a comprehensive study with data from more than 700 service companies, Rust and Huang showed that the use of automation in services does not automatically result in higher service quality or productivity (Rust and Huang, 2012).

Grönroos and Ojasalo (2004) therefore conclude, *"if improved productivity does not lead to better economic results, increasing productivity does not make sense"*. They claim that existing productivity models are manufacturing-based, assuming that production and consumption are separate processes and customers do not participate in the production process. As discussed above, this assumption is not true for many types of services. Whilst traditional productivity models usually are stated as the transformation of input resources into outputs with the quality of the output unchanged (a constant quality assumption), a changed set of inputs in a service operation may easily alter the perceived quality of the output. Thus, in the service context, productivity and perceived quality should not be understood as separate phenomena, but as two sides of the same coin (Grönroos and Ojasalo, 2004).

As service employees have a crucial influence on the perceived quality of the output, they are among the most important input resources in the production of services (Farquhar, 2004; Fitzsimmons and Fitzsimmons, 2008; Vargo and Lusch, 2008). They interact - often directly - with the customer, with little opportunity for management for intervention. Adequate assessment of employee´s performance, employee empowerment and a concern for training therefore becomes a necessity (Fitzsimmons and Fitzsimmons, 2008). Productivity of services is also frequently dependent on how

the relationship between service provider and customer develops. Several studies point out the relationship between perceived service quality and customer satisfaction leading to customer loyalty and retention (Farquhar, 2004; Hallowell, 1996). In this context, employee´s knowledge and skills are primary resources that are often undervalued in the process of value-creation and customer relationship (Vargo and Lusch, 2008). In a 2012 study about the effects of digitalization for retail banking, Bain & Company found that the performance and the quality of employees remain the essential distinguishing features in a competitive environment. Thus, more than 70 per cent of the study´s participants evaluated personal care and customer advisory services as important or even very important (Vater et al., 2012). This again emphasizes the importance of the organizations willingness to invest in careful employee selection on the one hand, and in continuous measurement and training activities on the other.

Due to varying levels of employee´s experience, education or social competencies and the varying levels of customer´s perception of output, it seems reasonable, that service employees have no unique production function to produce the same service. Therefore, when it comes to assess and measure employee´s performance, the measurement method should account for those variations. From a managerial point of view, the management of service employee´s performance should contribute to interactions with customers in a way that creates an optimal balance between perceived quality and internal efficiency (Grönroos and Ojasalo, 2004).

5.4 Banking Services

The financial sector comprises a bundle of institutions providing a bundle of services including: facilitating transactions (exchange of goods and services) in the economy, mobilizing savings (for which the outlets would otherwise be much more limited) or allocating capital funds (World Trade Organization 2015). In the following section the importance of the financial system and financial institutions for an economy is briefly outlined. Since the case study, which is illustrated in the scope of this thesis, was carried out among bank employees in Germany, the investigation holds a specific focus on the German financial sector. Finally, banking services are classified by adopting the classification schemes of Schmenner (1986) and Silvestro et al. (1992, 1999).

5.4.1 The role of financial institutions in an economy

The traditional role of financial institutions is to take deposits and to make loans. Those deposit taking and lending activities are commonly referred to as “commercial banking”. An increasing area of banking is “investment banking”, which deals with assisting companies or governments in raising debt and equity or providing advice on mergers and acquisitions and other corporate finance decisions. Since investment banks hardly deal with the general public (like individual savers or SME´s), they are subject to less regulation than commercial banks. Today, most large banks engage in both commercial and investment banking (Hull, 2012). In the onset of the GFC in 2008, new regulations – especially for investment banking – have been established at the EU and on international levels by tightening solvency and liquidity requirements.

However, not only because of the financial crisis but also because of a long-standing trend of rationalization, the total number of financial institutions keeps on declining in the EU (in 2014 by around 2,000 decreasing to a total of 7,267 institutions). This rationalization has particularly affected the bank´s branch network where a total of 34,150 branches were closed since 2009. In 2014, the total number of branches fell by 7,544 (3.6 per cent), from 211,861 to 204,317 bank branches in the EU. Certainly, the decline of bank branches has also affected the employment rate. Thus, the rate of staff employed in the European banking sector dropped continuously since 2008. In 2014 it decreased by 2.5 per cent to 2,889,320 employees working in banks in the EU (European Banking Federation, 2015).

5.4.2 The financial sector in Germany

The financial sector is of crucial importance for the German economy. It contributes by about 4 per cent to the GDP and employs about 1.2 million people in total (German Federal Statistical Office, 2015). The banking industry is a key component of this sector. Basically, the German banking system consists of two types of banks: The German central bank (“Bundesbank”) and commercial banks. The main task of the Bundesbank is to offer cash and central bank balances to the commercial banks. Commercial banks are operating in various legal forms. The business purpose of these banks is governed by the requirements of shareholders and often pursues the aim of maximizing profits. The legal name of a commercial bank is “financial institution”. In accordance with the German Banking Act (1998) a financial institution is a commercial organization, which conducts banking businesses. These transactions include the management of accounts, the management of deposits and securities and the accommodation of payments and

credits. Depending on the business purpose, the German commercial banking system differentiates between universal banks and specialized banks

One of the main characteristics of the German commercial banking system is its "three pillars model"[7], comprising private-owned commercial banks, public sector banks and cooperative banks (European Banking Federation, 2015). Unlike many Anglo-American countries, where a universal banking system was banned and has led to the development of a two-tier banking system, universal banks are the predominant form of organization in the German banking sector.

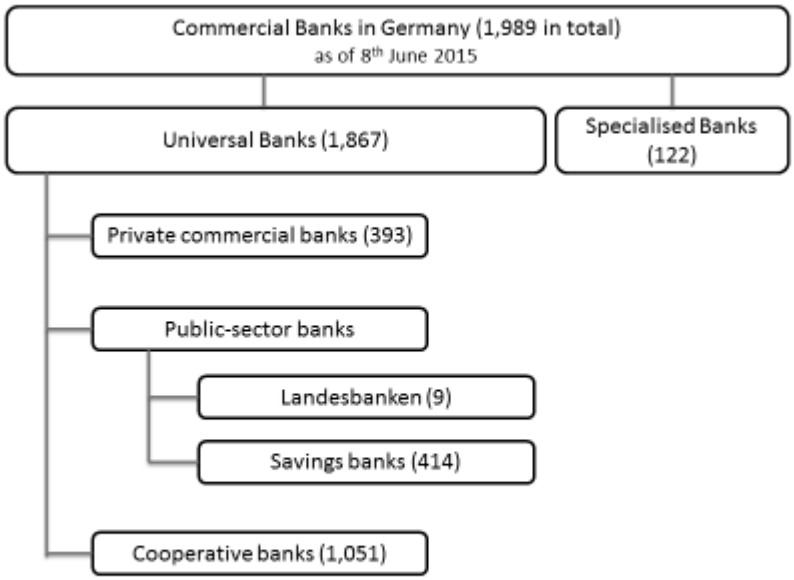


Figure 22: The Banking system in Germany (German Federal Bank, 2016)

The distribution of employees between the different types of commercial banks is as follows:

[7] These three pillars do not include specialiesed banks like building societies.

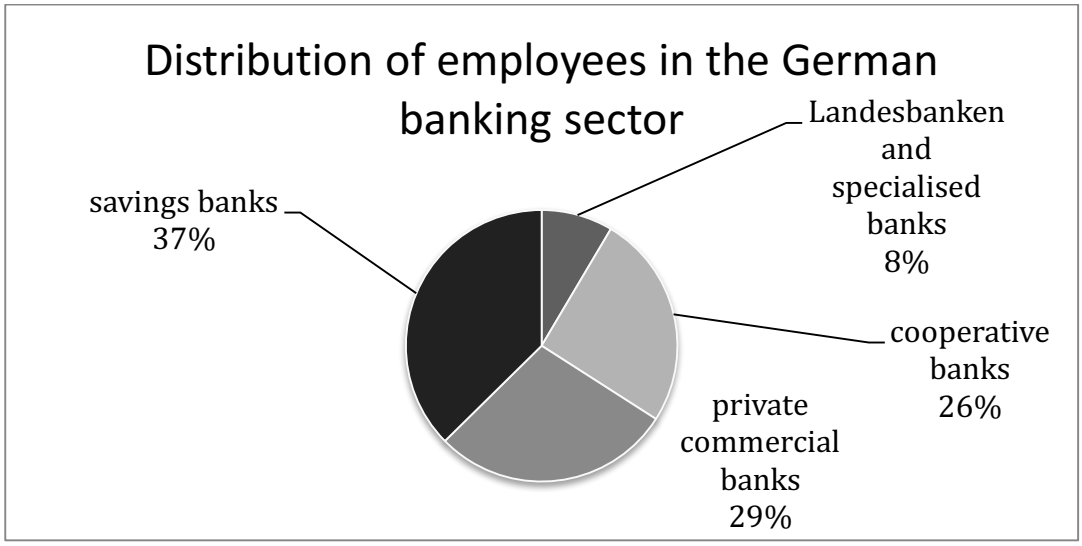


Figure 23: Distribution of employees in the German banking sector (Federal Association of German Banks, 2017)

The largest sector is represented by the private-owned commercial banks. Although there are currently only 393 private bank institutions, they accounting for about 39 per cent of total assets in the banking system. Some of them are operating worldwide, but there are also a lot of small banks with a rather regional focus (European Banking Federation, 2015). A major distinguishing feature of private banks is their competitive orientation. Contrary to public sector banks and cooperative banks, they not only compete with other banks in other sectors of the industry, but also among themselves.

The public banking sector comprises two different kinds of institutions: savings banks ("Sparkassen") and Landesbanken. In total, they account for about 28 per cent of bank assets. Savings banks are organized as public-law corporations with local governments as owners or guarantors. Thus, they have a very regional focus and usually limit their business to their local area.

Cooperative banks represent the third "pillar". Although they account for only 14 per cent of total bank assets, they represent 55 of institutions by number. A distinctive feature of cooperative banks is that they are owned by their members. Those members usually represent about half their customers and mainly consist of SMEs. Although, by virtue of their legal form, they have a mandate to support their members, they do also provide banking services to the general public. Like savings banks, they have a regional focus. Even though they are not legally bound to follow the regional principle, they usually do not compete with one another (European Banking Federation, 2015; German Federal Bank, 2016).

In line with the European trend, the total number of financial institutions in Germany has decreased within the last decades. Whilst the total number of banks was 4,582 in 1990, this number dropped sharply to 1,989 in 2015. Accordingly, a decline can be observed for bank branches, whose total number fell from 42,100 in 2007 to 33,914 in 2016. Thus, it comes as no surprise that the number of staff employed in the banking sector shrank by almost 10 per cent from 2007 to 2016. Interestingly enough, the number of customers increased by more than 15 per cent during the same period (Federal Association of German Banks, 2017). Although the number of customers has constantly risen over the last decade, market growth has stagnated. Thus, the competition between the financial institutions increases. Especially retail banking becomes more attractive due to less regulatory requirements. On the other hand, regulations on consumer protection and on documentation requirements were strengthened, which leads to higher costs. Additionally, customers feel less tied to their bank, the confidence in the sector as a whole has fallen rapidly. Banks therefore not only have to work on their efficiency, but to improve their services and align their core services more closely to customers' wishes (Pricewaterhouse Coopers, 2011). Again, employees are a key factor to achieve this purpose.

5.4.3 Classifying Banking Services

Since banking services are rather complex, consisting of several sub-services, it seems reasonable to classify banking services for further assessment. In service sector research, several classification schemes have been proposed over time. They, however, significantly diverge with regard to their concepts and procedures (Birkmeier et al., 2015). The most common classification schemes were illustrated in chapter 5.2.2. In the following, banking services are classified by applying the classification schemes of Schmenner (1986; 2004) and Silvestro et al. (J. Lee and Lee, 2014; Silvestro, 1999).

Schmenner did not apply his classification scheme to banking services in particular. However, according to the definitions of the four distinct services types, banking services may be classified as follows:

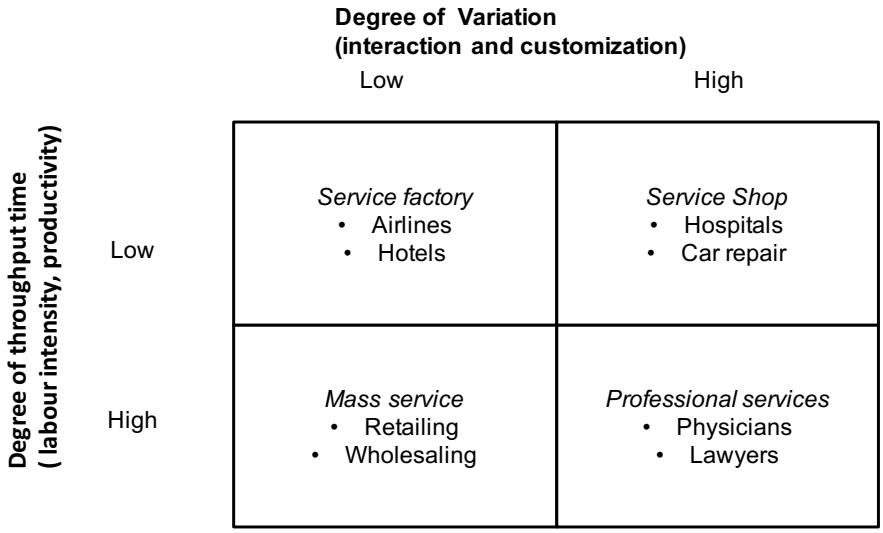


Figure 24: The service matrix by Schmenner (2004): positioning banking services

The classification scheme by Schmenner (2004) clearly shows that banking services consist of a variety of distinct service types that relate to the IHIP characteristics by varying degrees. “Professional services”, for instance, are of intangible nature. Standardization of this type of service is only possible to a certain degree. Although instruments like standardized checklist or contract templates may be used, the counselling process itself is very individual – if it was not, this type of banking service would soon become superfluous, as customers could get standardized financial advice elsewhere. Highly specialised financial counselling services also require the presence of the customer as a participant in the service process. Although the customer does not need to be physically present (consulting by phone or chat are would also be conceivable), the service cannot be produced in advance or in the customer´s absence.

Whilst “mass services” are also of rather intangible nature, they can be standardized in many cases. “Mass banking services” include services where customer´s expectations often are predictable and quite homogenous. In some cases, such as cash withdrawal from ATMs or online money transfer – mass banking services can be accomplished automatically without direct employee involvement. In other cases, such as the provision of foreign currencies, the service does not require the presence of the customer and can easily be produced in advance.

The service process classification scheme developed by Silvestro et al. in 1992 was modified later by Silvestro (1999) applying his model to banking services (Tinnilä, 2012):

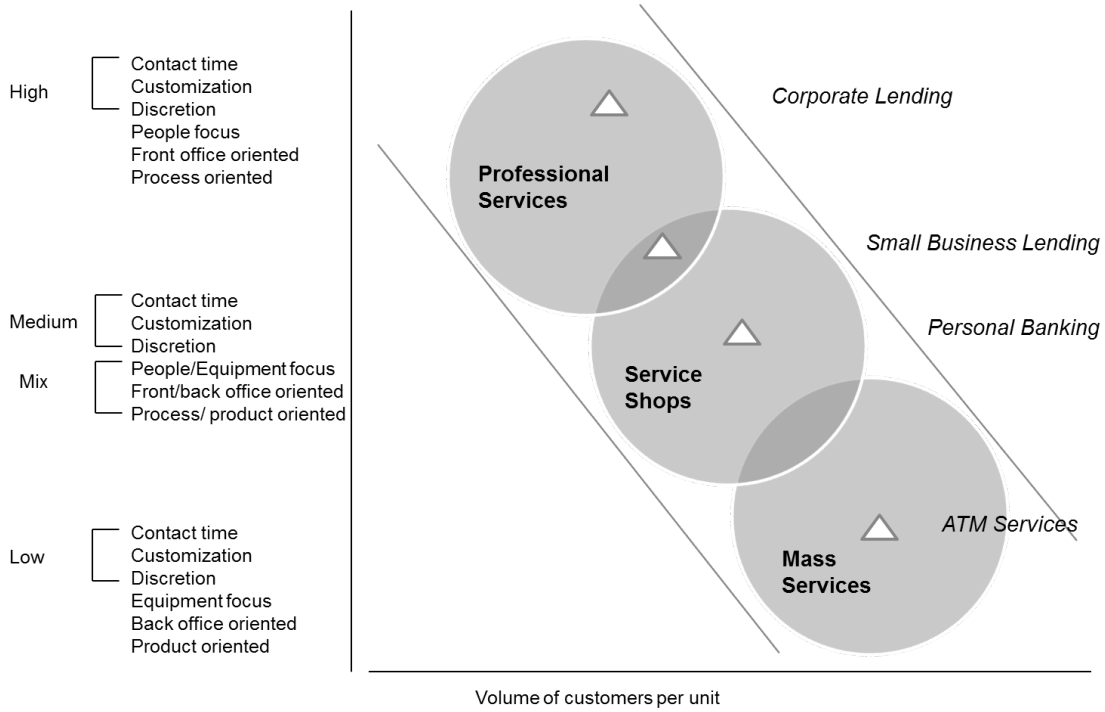


Figure 25: positioning of banking services (Silvestro, 1999)

Again, the classification shows that banking services consist of a variety of distinct services types that are positioned along the diagonal of the service process. According to Silvestro (1999), corporate loans, characterized by a relatively low volume of customers, customized service offerings and contact to highly skilled professionals, are a concise example of professional services. Mass services, at the other end of the diagonal are characterized by a high volume of customers, highly standardized offerings and a low level of contact. Additionally, those services are often highly automated (equipment rather than people ratio) such as ATM services. Service shops respectively, are characterized by medium customer volume, customization and contact levels. Personal banking and small business loans are concise examples for business shops.

Regardless of the underlying classification scheme, the varying manifestations of service characteristics in banking services impose challenges for management. In particular, the labour intensive services types “mass services” and “professional services” require differentiated activities in Human Resources Management.

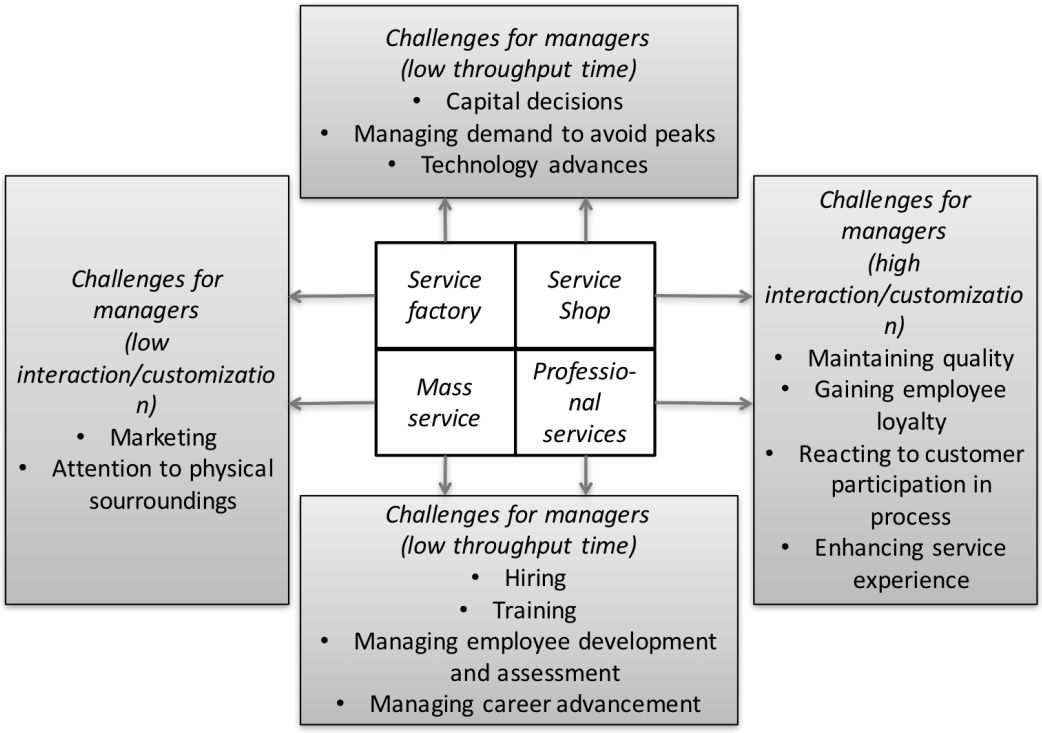


Figure 26: Challenges for service management (adopted from (Brandon-Jones et al., 2016; Schmenner, 1986))

Since both service types are labour intensive, they require adequate hiring and training processes. The customized nature of professional services requires high discretion in meeting customer requirements. Customers often take actively part in the service process and may even change specifications during the course of the process. Thus, identifying training needs for contact employees in professional services is an essential task of leadership (Benkenstein et al., 2017). Furthermore, customers often build long-term relationships with individual members of staff. Thus, employees often need to be highly qualified with skills that are difficult to acquire. In addition, the way a professional service is produced is likely to differ from employee to employee. In this respect, generating employee loyalty, keeping retention rates high and managing career advancement are likely to be key concerns for HRM. Hence, adequate employee assessment and development plays a major role to remain competitiveness. To a certain degree, this is also true for service shop services.

Mass services, on the other hand, tend to be highly reproducible, which may require employees to be tolerant of repetition. Contrary to professional services, part-time and less skilled employees can well be used to increase flexibility. Since mass services are more standardized, measurement of performance is easier to accomplish (Silvestro, 1999) and production processes are rather similar.

6 Research methodology and conceptual framework

In the following chapter, the conceptual framework and the research methodology employed for addressing the research objectives are presented. Initially, the findings from the literature review are brought together to inform the construction of a conceptual framework. Subsequently, the research questions guiding this research are located within the conceptual framework.

The chapter continues by discussing the dominant research paradigms along with offering the researcher's philosophical stance. From this, the rationale of the chosen research strategy and design for carrying out this study are outlined. Further, the methods adopted for data collection and analysis are illustrated, reflecting on their strengths and weaknesses and justifying their application. Closing, the researcher draws conclusions on the research project´s reliability and validity and also offers a discussion on the major limitations of her research.

6.1 The Conceptual framework

In this chapter the findings from the reviewed literature are summarized. The literature summary subsequently informs the construction of a conceptual framework and by this the location of the research questions. Further, the conceptual framework provides the basis for the choice of research methodology.

6.1.1 Synthesis of literature review

Although DEA has become a well-known method for evaluating the performance of organizations or organizational units since it was first proposed in 1978 (Emrouznejad and Yang, 2017) there is very little experience on applying DEA to evaluate employee performance. Further, no study provides empirical data about employees´ responses towards the DEA procedure. Moreover, there is a lack of information about whether DEA´s results are helpful to inform administrative and developmental HRM actions. Both purposes play an important role for the employment and development of people and their individual resources to support the organization achieving its objectives (Chartered Institute of Personnel and Development, 2016a). Although there has been some major criticism considering the role of HRM in general including the question how HRM and HR practices may have contributed to the recent global financial crisis, there is strong evidence in research that evaluating employees´ performance is a worthwhile process, when it is done "right". Thus, it is crucial to understand the specific requirements under

which performance evaluation may deliver the promised results and how the whole procedure should be conducted properly (Chartered Institute of Personnel and Development, 2016a).

In this respect, literature holds a variety of recommendations on what constitutes a “good” or “adequate” performance evaluation. Thus, perceived fairness of the process and the results of the evaluation are cited as the most important requirement by most studies (Sudin, 2011). However, it should be noted that the success and the achievement of the intended purposes and benefits of performance evaluation not only depend on a proper design but also on management style, the power relationships between stakeholders and other contextual environmental factors. Further, the review of literature on service management clearly showed that some services place advanced requirements to an evaluation approach.

Especially services that offer a high degree of interaction and customer integration strongly rely on the relationship between employees and their clients. Furthermore, profitability of these services and performance of employees is difficult to ascertain due to variety and choice that is often provided to the customer. Therefore, employee´s knowledge and skills are primary resources for organizational success. To produce a service that meets the customers´ demands for quality, employees may have varying production processes. Further, the production of a service may be influenced by environmental factors that are beyond the employees´ control (e.g. the location of the workplace). Thus, an adequate performance evaluation method needs to consider all these requirements.

6.1.2 Construction of the Conceptual framework

The comprehensive literature review and the subsequent findings build the basis for the conceptual framework of this thesis. The conceptual framework can be viewed as a structure for clarifying research issues and organizing the content of the research by linking concepts and findings from the literature review. Thus, it illustrates the theoretical overview of intended research and provides a set of reference points from which to locate the research questions (Leshem and Trafford, 2007) .

Synthesizing from the literature review, the figure below illustrated the conceptual framework of this research. It also indicates the reference points for the research questions.

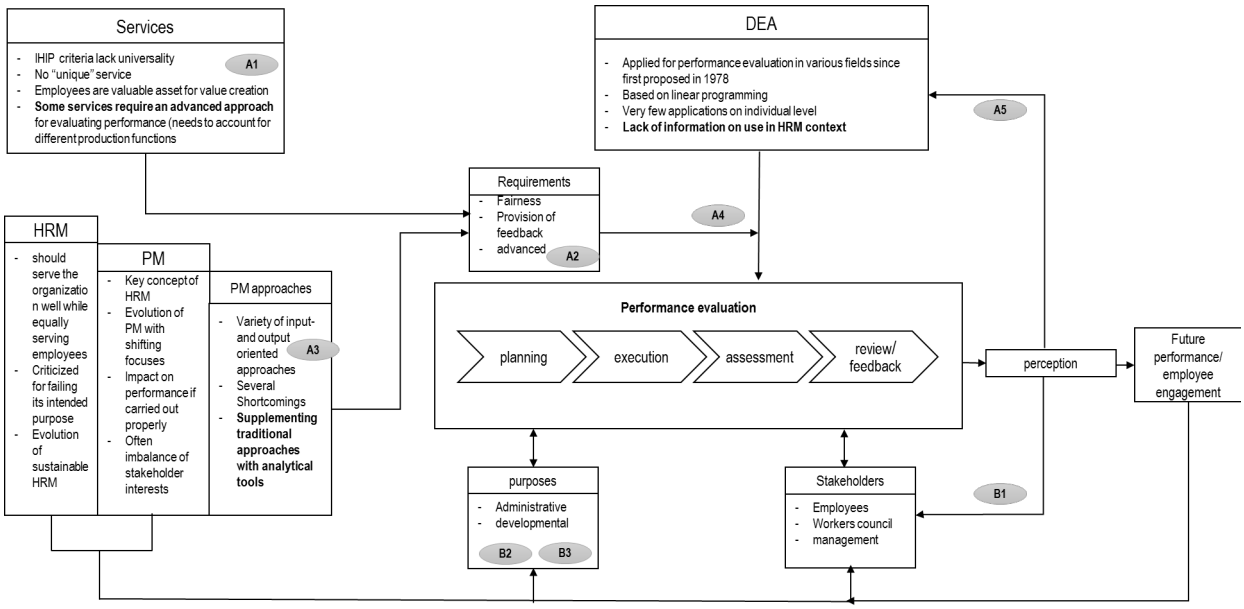


Figure 27: Conceptual Framework

As outlined in chapter 1.4.1, the major aim of this research is to investigate whether the Data Envelopment Analysis is a suitable approach to evaluate employee performance in the service sector. Thus, the main research objectives are to investigate:

A: DEA´s technical suitability for evaluating employees´ performance in the service sector

B: DEA´s organizational suitability for evaluating employees´ performance in the service sector

From the research objectives, several subsidiary research questions are drawn (see chapter 1.4.). Addressing the research question whether the evaluation of employee performance in the service sector actually requires an advanced approach (A1) the literature review on service sector research reveals that there is no "unique" service but that there are different service types placing distinct requirements to performance evaluation (Grönroos and Voima, 2013). Thus, "mass services" are in many ways comparable to the production of goods. Therefore, traditional manufacturing based productivity models and evaluation methods may apply. On the contrary, "professional services" and "service shops" are service types that place different requirements to the evaluation of employee performance. Investigating research question A2 ("What are (…) the requirements for an advanced approach?") the review shows that an evaluation approach for those particular service types requires taking several input and output factors into account. In addition, due to varying production processes, the evaluation approach should be able to discriminate between different production processes and to

give detailed information about relative distance between employees (Vargo and Lusch, 2008). The analysis also indicates that for those service types, employees often are a source of unique skills and knowledge and therefore are a key driver for competitive advantage.

Considering the importance of skilled and motivated employees for service organizations, the review of HRM literature shows, that HRM activities are essential for the organization in order to employ and develop people (Armstrong, 2018). However, the concept of HRM has been facing criticism for neglecting its intended purpose to serve both sides, employees and managers or shareholders (Keeble-Ramsay and Armitage, 2015). To become mutual beneficial, the approach of sustainable HRM has emerged as an alternative to HPWS. With focus on long term and sustaining development, the concept holds an increased focus on employees as important stakeholders to the organization (Kramar, 2014). For developing employees, PM is a key function within the concept of HRM which also has evolved over time. Whilst the focus of PM was on accountability and on administrative purposes, a developmental orientation of PM became more popular during the last decade (Cappelli and Tavis, 2016). More recently, the trend of Business Analytics became popular in the field of PM by supplementing traditional approaches with sophisticated and analytical decision-making tools (Schläfke et al., 2012). Although the impact of PM on performance is still questioned by some researchers, there is a growing body of evidence that PM can be a valuable process if carried out properly (Chartered Institute of Personnel and Development, 2016a). Among the most cited general requirements for a proper performance evaluation (research question A2) is that a performance evaluation primarily needs to be considered fair by the employees. In addition, for managing career advancement, detailed feedback information for improvement needs to be provided by the evaluation method (Chartered Institute of Personnel and Development, 2016a; Kondrasuk, 2012; Sudin, 2011).

By investigating the question what the pros and cons of traditional approaches for performance evaluation are (A3), it was found that, besides several strengths, traditional approaches show several shortcomings. Thus, traditional absolute approaches are often perceived as biased and are not able to provide useful improvement information. Comparative approaches fail to integrate and aggregate several input and output factors or to account. Also most approaches fail to discriminate between different production processes or to account for uncontrollable factors.

Considering that some service types demand advanced requirements for evaluating employee performance and that most traditional PM approaches fail to account for those requirements, DEA may be a promising approach to addresses some of the

shortcomings of traditional approaches. As the review of literature reveals, DEA has been applied in various organizational context for evaluating performance since it was first proposed in 1978. However, until today there have been only very few applications on employee level. Moreover, DEA´s use in the context of HRM has not been investigated so far. Moreover, there is no information on how the method was perceived by employees.

In this thesis, DEA is applied for evaluating the performance of service employees in the banking industry according to the PM cycle (Armstrong and Taylor, 2014). To investigate whether DEA is able to meet the advanced requirements (research question A4) and provides accurate and reasonable results on employee level (research question A5) the method and results need to be analyzed and thereafter reviewed by all main stakeholders. The review should also examine how the method is perceived by all stakeholders (research question B1) and whether it accounts for both administrative and developmental purposes (B2). This is particularly important considering the lack of empirical data referring to employees´ perceptions of performance evaluation. Since DEA is applied in the banking industry, it should also be studied if the obtained results are a suitable basis for determining performance targets and bonus payments (B3).

Thus, learning more about DEA´s suitability for evaluating employee performance and addressing some recurring issues considering performance evaluation in general may add some value to the scientific discussion.

6.2 Research Design and Methodology

The previous section provided a synthesis of the literature review and pointed out the major findings from literature and the existing gap in current research. Those findings were connected in a conceptual framework that highlights the key objectives of this research and located the research questions. While the research questions A1 to A3 have already been addressed by the comprehensive literature analyses, the investigation of research questions A4 and B1 to B3 requires an empirical investigation. To adequately address those research questions, the research design and the underlying research methods need to be selected carefully. Thus, the following chapters illustrate the adoption of a research philosophy and describe the choice of the research methodology and the research methods applied.

6.2.1 Research Philosophy

To examine and classify the process of systematic investigation that constitutes research, a variety of philosophical and theoretical perspectives exist. Thus, the first steps in a research process is to acknowledge the underlying research philosophy of one´s research. According to Kuhn (1969) a research philosophy - or paradigm – can be regarded as an “accepted model or pattern” or a deeper philosophical position that relates to the nature of social phenomena. In this context, a research philosophy can be interpreted as how the researcher thinks about the world including beliefs about her effort to create knowledge (Morgan, 2007). According to Creswell and Clark (2007), researchers approach their studies with a certain worldview that relies on a basic set of assumptions that guide their inquiries. Subsequently, these assumptions underpin the research strategy and methods and therefore have an impact on the way the research is conducted and the findings are analyzed (Saunders et al., 2012).

In the following sections, four major paradigms (positivism, constructivism, pragmatism, critical realism) along with their respective philosophical assumptions will be discussed. The discussion of the distinct paradigms will, however, not address the so called “paradigm wars” (Feilzer, 2010). Rather, the discussion will build the basis to select and justify the underlying paradigm of this research.

To characterize philosophical positions aspects on ontology, epistemology and methodology are usually addressed. *Ontology* describes the nature of reality. This in particular concerns assumptions about the way the world operates (Saunders et al., 2012). There are two different aspects of ontology: Objectivism that portrays the view that the social world is something external to the social actors concerned with their existence, and Subjectivism that holds the position that the social world is created through the perceptions and actions of social actors (Bryman, 2015; Saunders et al., 2012). *Epistemology* addresses the question of what is regarded as appropriate knowledge about the social world and how this knowledge is developed. It establishes the relationship between researcher and reality. Again, there are two distinct aspects to epistemology: the positivist philosophy to the development of knowledge where reality is represented by objects that are considered to be “real”, and the constructivist philosophy that rather relies of the perceptions and attitudes of social actors (Saunders et al., 2012). The *Methodology* deals with the question how the researcher goes best to investigate the research subject. The choice of methodology therefore has an impact on the choice of research methods (Bryman, 2015).

6.2.1.1 Research Paradigms

When the work of Kuhn began to gain momentum (1969), there was no commonly agreed upon label for the dominant paradigm that characterized social science research methodology up to that point. For an existing paradigm to lack a label and a clear characterization of its content this may not be unusual, until it is called into question by a set of challenges (Morgan, 2007). In the case of research paradigms, the existing dominant approach that was later labelled as "positivism" was questioned at the level of fundamental assumptions by an alternative paradigm that today is acknowledged as "constructivism" (Guba and Lincoln, 1994a).

Positivism

Historically, there has been a heavy emphasis on quantification in research. Hence, positivism that dominated the "received view" in physical and social sciences for more than 400 years, focuses on efforts to verify a hypothesis by stating quantitative propositions (Guba and Lincoln, 1994a). Researchers that adopt a positivist view believe that research can be conducted value-free and that only observable phenomena can genuinely be warranted as knowledge (Bryman, 2015). Hence, the positivist researcher considers things to be real if they are objectively measurable either through hypotheses testing or cause-effect relationships (e.g. in case of performance evaluation the outcomes that derive from the evaluation). In this sense, positivism contributes to the representation of human beings as things or mere objects of study (Harvey, 1990). The goal of positivist research is to discover general laws to describe constant relationships between variables. A general claim of positivist research is that scientific knowledge is utterly objective and that only objective knowledge is valid (Mertens, 2014).

Constructivism

In contrast to positivism, constructivism (sometimes referred to as "interpretivism") is an alternative paradigm whose assumptions rely on the idea that there is no single objective reality and that therefore subjective methods of inquiry are necessary (Feilzer, 2010). Researchers that adopt a constructivist view belief that the object of study in social science is different from that of the natural science. Therefore, scientific models are not suitable to study the social world. Rather, research should attempt the interpretive understanding of social action and subsequently arrive at a causal explanation (Bryman, 2015). One basic assumption of the constructivist paradigm is that reality is socially constructed and researchers therefore should gain interpretive understanding of those

involved in social action. In contrast to the positivist philosophy, constructivism emphasizes the difference of conducting research among mere objects and among people. People therefore are labelled as “social actors”. This metaphor plays an important role considering the constructivist worldview. The term “social actor” suggests that people play a role which they interpret in a certain way and which is also interpreted by others in accordance with their own set of meanings. In consequence, this interpretation may also lead to adjustments of their own meanings and actions. (Saunders et al., 2012).

An important issue to the constructivist philosophy is that the researcher has to adopt an empathic stance. In the case of performance evaluation, a constructivist approach would consider the need of a deeper understanding of the performance evaluation procedure by interacting with employees so to exemplify its impact on employee performance. In consequence, constructivist research is defined through the interaction of researcher and research objects which mostly requires qualitative research methods (Guba and Lincoln, 1994a).

Critical Realism

Critical realism is a paradigm that relates to the position that reality is a result of social conditioning. Hence, researchers only understand the nature of social phenomena if they understand the social structures that gave rise to those phenomena (Bryman, 2015). Researchers adopting a critical realism philosophy often focus on issues that are difficult to understand, such as radical changes within the economic and business environments. They share the belief that structural, contextual and ideological factors can significantly influence individual perceptions of reality. As individual accounts are important in the generation of theory, critical realism rests on the assumption that the accounts of research participants are valid social scientific data and that each of these accounts has the capacity to change the researchers understanding of what is being studied (Egbo, 2005). By acknowledging the socially-embedded and imperfect nature of scientific inquiry as well as the existence of knowledge independent of humans, critical realism holds a position between positivism and constructivism (Clark, 2008).

If interpreted properly, the gathered information on social phenomena can lead to social transformation (Egbo, 2005). Thus, researchers that adopt a critical reality philosophy often address politics in research in an effort to bring about social transformation (Mertens, 2014). Therefore, O’Donnell et al. (2006) claim that HRM research should adopt a more critical stance towards considerations related to employees’ problems and a power imbalance in HRM. Moreover, Sambrook (2004) suggested that a critical

approach allows different voices to be heard over the same topic so that multiple realities of it are revealed. However, some researchers point out, that the philosophy of critical realism is rather oriented toward “problem-posing” than “problem-solving” (Elliott and Turnbull, 2004).

Pragmatism

A pragmatic approach to research accepts that no single point of view can ever give the entire picture and that there may be multiple realities which are open to empirical investigation. It orients itself towards solving practical problems in the “real” world and therefore judges the meaning of an idea or research finding by its practical consequences (Feilzer, 2010; Saunders et al., 2012). In this sense, pragmatism sidesteps the forced choice dichotomy between positivism and constructivism (Creswell and Clark, 2007). Pragmatist researchers tend to favor the integration of both qualitative and quantitative methods claiming that they are both useful for empirical inquiry since one approach´s weaknesses can be diminished by the respective strengths of the other. Furthermore, results from a qualitative approach can serve as inputs to a quantitative approach, and vice versa (Morgan, 2007). Since the pragmatic paradigm emphasizes that social phenomena may consist of different layers or views, they often rely on mixed methods research avoiding being constrained by a single, monolithic method (Feilzer, 2010; Mertens, 2014). Morgan (2007) argues that the forced dichotomy between subjective and objective research is an artificial one since there may be no “complete objectivity” but there is also no such thing as “complete subjectivity”. Therefore, the possibility of being either completely subjective or objective is only a theoretical discussion. Hence, a pragmatic approach is also an intersubjective approach offering a practical solution to the discussion concerning the use of quantitative or qualitative methods (Mertens, 2014). Pragmatic researchers choose methods or method combinations of which they think they may work best for answering their research questions (Johnson and Onwuegbuzie, 2004).

To allow a well-reasoned decision of the underpinning philosophy for this research, the characteristics of the discussed research paradigms are summarized in the table below.

Paradigm	Ontology "What is reality"	Epistemology "What is acceptable knowledge?"	Methodology "How to go to find out?"	Methods "What techniques are applied to find out?"
Positivism	There is only one reality that is objective and independent from social actors,	Reality can be measured. Only observable phenomena provide credible data and facts.	Experimental research, survey research, etc.	Usually quantitative including: Statistical Analysis, Questionnaires
Constructivism	There are multiple realities. Reality is subjective and created by social actors.	Reality is subjective and needs to be interpreted. The focus is on details of the situation and on underlying meanings of events and activities.	In depth analysis, Ethnography, Grounded Theory, Discourse Analysis, etc.	Usually qualitative including: Qualitative Interviews, focus groups, Narrative Analysis
Critical Realism	Reality is socially constructed and under constant influence.	Phenomena create sensations that are open to misinterpretation. Reality and knowledge is socially constructed.	Critical discourse analysis, critical ethnography analysis, etc.	Quantitative or Qualitative, Method must fit the subject matter
Pragmatism	There are multiple realities that are constantly renegotiated and debated. Reality is interpreted in the light of its usefulness to answer the research questions.	Either or both observable phenomena and subjective meanings can provide acceptable knowledge dependent on the aim of the research and the research question.	Multiple or mixed methods, Triangulation, etc.	Usually a combination of Qualitative and Quantitative methods

Table 9: Comparison of four research paradigms *(Guba and Lincoln, 1994b; Mertens, 2014; Saunders et al., 2012)*

6.2.1.2 Justification for adopting the pragmatic research philosophy

The aim of this research is to investigate the technical and organizational suitability of Data Envelopment Analysis for evaluating employee performance in the service sector. Considering the analysis of DEA´s technical suitability, a positivist approach seems reasonable since the collection of objective, quantitative data seems to be the best methodology testing the validity of results. Further, to ensure objective results, the researcher should be external to the process. However, this philosophy falls short when it comes to evaluate DEA´s organizational suitability or more precisely its perception by the stakeholders of the evaluation process. As the literature review has shown, a performance evaluation method needs to provide accurate feedback information and needs to be perceived as fair by the employees (Chartered Institute of Personnel and Development, 2016a). In consequence, to answer the research questions considering DEA´s organizational suitability, the perceptions (realities) of all major stakeholders need to be considered. First, this implies that there may not be one single and absolute truth to this question. Second, the researcher cannot be separated from the object of research since the object (application of DEA) is embedded in an organizational context and needs to be analyzed against this background. Thus, there needs to be a certain involvement of the researcher in terms of interacting with stakeholders and analyzing and interpreting their views. Therefore, positivism does not provide a reasonable position for this research.

Constructivism, as an alternative paradigm, advocates that it is necessary for the researcher to be actively linked to the object of investigation and engage in interactions with the respondents (Guba and Lincoln, 1994b). Furthermore, the constructivist researchers rejects the idea that there is only one truth or reality (Mertens, 2014). However, Foley (1995) highlighted that constructivism often lacks structure. Moreover, constructivist researchers are skeptical about being able to actually capture reality at all and often becomes a “passionate participant” of its research (Perry et al., 1999). The investigation into the suitability of a complex method like DEA along with the consideration of the advanced requirements to evaluating employee performance call for a more comprehensive analysis than a rather unstructured interaction with stakeholders. Moreover, the assumption that all perceptions of stakeholders are constructs of reality makes it difficult to gain a general understanding of DEA´ suitability for the given purpose. Therefore, the constructivist paradigm also is not a “good fit” for this research.

Critical realism adopts views from both positivism and constructivism (Perry et al., 1999) and therefore is open to qualitative and quantitative methods of inquiry. It poses questions of “why” and “how” so to examine specific phenomena. Moreover, it takes a

critical stance towards power conflicts as observed in HRM (see chapter 3.1.3). Thus, a critical researcher could interview and pose questions to certain groups of stakeholders (e.g. HR managers for instance) which are considered crucial from a disadvantaged groups' perspective. Although employees are not a disadvantaged group in the proper sense, they are subordinates in a hierarchy of power relations and therefore may not be regarded as "equal" by other stakeholders. Thus, critical researchers are interested in the underlying observable and non-observable structures and mechanisms of these phenomena. According to the critical realist ontology, social phenomena are made possible by the presence of humans but are deemed to be external to individuals and have existence whether this being recognized by individuals or not (Archer et al., 2013). Therefore, critical realism distinguishes between actual, real and empirical world views. Epistemologically, the knowledge that is gained by critical research is real but also incomplete or fallible since there is one reality but due to different perceptions it needs to be captured by triangulation (Perry et al., 1999). Considering its openness to methods and its critical approach towards power imbalances, critical realism may be a reasonable philosophy to adopt for this research. However, the assumption of this research is that different stakeholders to the performance evaluation may experience different realities and the investigation should account for those differences. Further, this research aims at contributing to the recurring problem of measuring employee performance adequately. Thus, it is oriented towards solving a practical problem rather than to posing a problem.

From this stance it seems reasonable that this research follows a pragmatic research philosophy applying a mixed-methods approach in an iterative procedure. In agreement with the first research objective, it is necessary to apply DEA in an organizational context and obtain objective, observable and quantifiable data. To ensure objectivity, the researcher is not involved in the process of gathering the data on employee performance but relies on archival data provided by the organization. In agreement with the second research objective of this study, it is necessary to understand stakeholders' perceptions and personal perspectives on the suitability of the proposed DEA approach. Therefore, it is necessary for the researcher to become actively involved in the research and interact with the respondents and reflect their distinct views. Thus, methods of both quantitative and qualitative nature will be combined that allow reliable and relevant data to be collected in order to solve a rather practical problem.

6.2.2 Research Design

As the logical sequence that connects the empirical data to the study´s research aim and research questions the research design provides a logical path for structuring the

research actions. Thus, the following chapter introduces the research methodology and the underlying research strategy of this study. It provides a clarification for the use of a mixed-methods design and introduces the case study as the underlying research strategy. Finally, the methods for data collection and data analyses are presented.

6.2.2.1 Research approach

An important decision considering the design of the research project is which approach of considering the relationship between theory and social research is adopted. In general, there are two approaches of reasoning: an inductive or a deductive approach (Saunders et al., 2012). According to Bryman (2015), deductive research represents the most common view. Thus, the researcher deduces a hypothesis from a particular theoretical domain. This hypothesis then is subject to empirical scrutiny and is translated into researchable entities. Considering the direction of research actions, deductive research moves from generating a theory to collecting empirical evidence in order to test the theory through a series of propositions. One major characteristic of deduction is generalization from the general to the specific. In consequence, deductive research is the dominant research approach in natural science (Saunders et al., 2012).

Inductive research, on the other hand, aims to generalize from the specific to the general. With an inductive stance in research, theory is the outcome of research by drawing generalizable interferences out of empirical evidence (Bryman, 2015). The inductive approach emerged as an alternative to deduction. Critics of the deductive approach claimed that deduction enabled a cause-effect link to be made between particular variables without understanding the way in which humans interpret their social world. Thus, a major advantage of the inductive approach that it enables the researcher to gain a deeper understanding of social phenomena, explore the phenomena and identify themes and patterns to create a general theory (Saunders et al., 2012).

In recent years, there has been some criticism considering the rigor of the approaches. As noted in Saunders et al. (2012), Van Maanen et al. (2007) point out that some theories account for what is observed better than others and that “surprising facts” may occur at any stage in the process. Therefore, it may become necessary for the researcher to move from theory to data (deductive) and from data to theory (inductive) within the research project. From this criticism, abduction emerged. In general, abduction begins with observing a “surprising fact” from a known theory. It continues with data collection to explore this fact or phenomena, locate themes or patterns and place them in a conceptual framework. Again, these themes are tested through subsequent data collection and so forth (Saunders et al., 2012).

For this research, a recurring problem, namely the lack of an adequate method for evaluating employee performance in the service sector, was identified from literature. Although there is a wealth of literature on both service sector research and human resources management, this seemed to be a topic of special interest. Along with the observation that DEA, despite its long history as a method for evaluating performance, has hardly been used on individual level, this makes this topic a “surprising fact” in the sense of abductive research. From the literature, themes and patterns were derived and were brought together in a conceptual framework. Subsequently, the research questions were located within this framework. To address several research questions, empirical data need to be collected in a rather deductive process. However, the findings from this data collection process shall not built or generalize the theory on HRM or performance evaluation. Rather, the findings shall inform or enhance the existing theory on performance evaluation by assessing the organizational strengths and weaknesses of an innovative method to performance evaluation on individual level. In conclusion, this research project uses known premises from HRM and service sector research theory to generate testable conclusions and so to enhance or modify existing theory. Hence, this research project requires an abductive approach drawing on both inductive and deductive elements.

6.2.2.2 The nature of research

The choice of the research methodology and the respective methods for data collection mainly depend on the nature of the research problem. Thus, whether the chosen methodology is suitable to tackle the research problem derives from the nature of the social phenomena that is to be investigated (Noor, 2008). The purposes of research projects can generally be classified into three categories: exploratory, descriptive and explanatory research. A single study may have multiple purposes so that combinations can occur (Sue and Ritter, 2011).

The nature of exploratory studies is to gain deeper insights on a topic of interest or to clarify the researchers understanding of a problem (Saunders et al., 2012). Often, exploratory research helps to identify or formulate a research problem or a hypothesis (Sue and Ritter, 2011). The way exploratory research is conducted is manifold starting from comprehensive literature review over in-depth individual interviews to conducting focus group interviews (Saunders et al., 2012). Descriptive research, on the other hand, aims at describing characteristics of people, situations or events. In contrast to exploratory research it is more structured and tends to have more guidelines. Therefore, the researcher needs to have a clear picture of the phenomena she wants to describe.

Descriptive research may employ both quantitative or qualitative methods (Sue and Ritter, 2011). When a causal relationship is investigated, the research purpose usually is of explanatory nature. Hence this kind of research aims at studying situations or problems in order to explain relationships between variables. Explanatory studies often have a research hypothesis that specifies the nature and direction of the relationship. Hence, explanatory research mainly employs quantitative methods including sampling and requires statistical tests to establish the validity of the relationship (Sue and Ritter, 2011).

The research objectives of this study are twofold. The investigation of DEA´s technical suitability to evaluate employee performance is of rather exploratory nature. Thus, one major goal of this strand of research is to gain a deeper understanding of DEA´s application in the HRM context. Moreover, it tackles the issue of understanding the advanced requirements for performance evaluation in the service sector. For addressing the research questions A1 to A3, a comprehensive literature review was conducted, which is an often used method in exploratory research (Sue and Ritter, 2011). Also, the question whether DEA produces reasonable results on individual level (A4) is rather exploratory.

In contrast, the investigation of DEA´s organizational suitability requires us to study the perception of DEA´s results by the stakeholders of the performance evaluation. The organizational suitability can be assessed against clear criteria that are derived from the literature review. Thus, the investigation relies on the description of peoples´ perceptions and is rather structured. Besides the descriptive nature of this research objective, a part of the investigation is also explanatory since the study seeks to identify cause-effect relationships that link the requirements of a performance evaluation to its actual perception by employees. In conclusion, this research combines exploratory, descriptive and explanatory research and therefore requires a research methodology that accounts for the respective requirements of those research categories.

6.2.2.3 Research Methodology: The mixed-methods approach

To investigate the technical and organizational suitability of DEA, empirical evidence needs to be gained. Thus, an appropriate research methodology needs to be selected. In general, a researcher has the basic methodological choice to opt for a mono method using a single data collection technique or opting for a multiple methods design by using more than one data collection technique to answer the research questions. The multiple methods design, in turn, can be classified into multimethod research and mixed-methods research. While the first uses multiple methods for data collection, these methods are

restricted within either a pure quantitative or a pure qualitative design. A mixed-methods research, on the other hand, combines qualitative and quantitative data collection methods and their corresponding analytical procedures (Saunders et al., 2012).

As already pointed out in the previous chapter, the investigation into DEA´s technical suitability is of rather exploratory nature and requires merely quantitative data analysis techniques. To evaluate DEA´s organizational suitability the perceptions of the main stakeholders of the performance evaluation, namely the organization´s managers, the workers´ council and employees, are to be assessed. Again, this assessment requires both qualitative and quantitative methods. The assessment of managers´ perceptions will inform the reflection of several research questions considering DEA´s organizational suitability. Thus, this investigation requires direct interaction with the respondents and merely open questions, which indicates the use of qualitative methods. One major aim of this research is to include the employees´ voice into the assessment of an approach for evaluating employee performance. Thus, empirical evidence on employees´ perception of the approach and information on cause-effect relationships on this perception shall be provided. Hence, the explanatory nature of this research question requires a rather qualitative research. Considering the distinct purposes of the research objectives and their corresponding nature, a mixed-methods approach seems a appropriate choice for this study.

This choice is also in coherence with the underlying research philosophy of pragmatism. Thus, mixed-methods research is often attributed to the pragmatic approach since pragmatist researchers agree that research methods should be mixed in a way that offer the best opportunity to answer the research question. This way, mixed-methods research attempts to fit together the insights provided by qualitative and quantitative methods into a workable solution (Johnson and Onwuegbuzie, 2004). The choice of the mixed-methods approach also warrants a further discussion about the purpose of combining qualitative and quantitative methods. Saunders et al. (2012) list ten distinct purposes, Bryman (2015) lists sixteen purposes respectively, that may account as reasons for using a mixed-methods design. Among them are initiation (i.e., discover contradictions that emerge when findings from the two analytical strands are compared. This again can lead the researcher to reframe the research question), diversity (to allow for a greater diversity of views to be presented in the study) and triangulation. When applying a mixed-methods approach for the purpose of triangulation, the researcher usually attempts to combine qualitative and quantitative research in order to ascertain that the findings from one method mutually corroborate with the findings from the other method (Saunders et al., 2012). In turn, the purpose of the mixed-methods design leads to various ways

mixed-method research can be conducted. In particular, the specific mixed methods approaches are defined by the order quantitative and qualitative methods are applied. Basically, the methods can be applied simultaneously or sequentially. If carried out simultaneously the distinct types of data are collected and analyzed in parallel form. If mixed-methods research is carried out sequentially, one type of data provides a basis for collection of another type of data (Mertens, 2014).

For this study, the mixed-methods approach is applied to gain different perspectives and subsequently cross-check the findings in order to assess DEA´s suitability for performance evaluation on individual level. Hence, the main purpose for choosing a mixed-methods approach is triangulation. To obtain empirical evidence, DEA first needs to be applied in a practical context. Therefore, the proposed method is applied to measure the performance of 40 service employees in a German Cooperative Bank. The application is based on archival data from 2012 that was provided as linked-anonymous data by the bank. In a quantitative analysis, the results of the application are examined to assess DEA´s technical suitability and to gain a deeper understanding of the DEA application on individual level. This assessment in turn builds the basis for assessing its organizational suitability by leading to formulate questions for data collection and data analysis. Subsequently, a qualitative data collection and analysis among the managers of the organization is carried out and might in turn provide valuable connecting points for the assessment by the employees. In conclusion, the data collection and analysis of this study is carried out sequentially.

The figure below presents an overview on the sequential mixed-methods research methodology of this study.

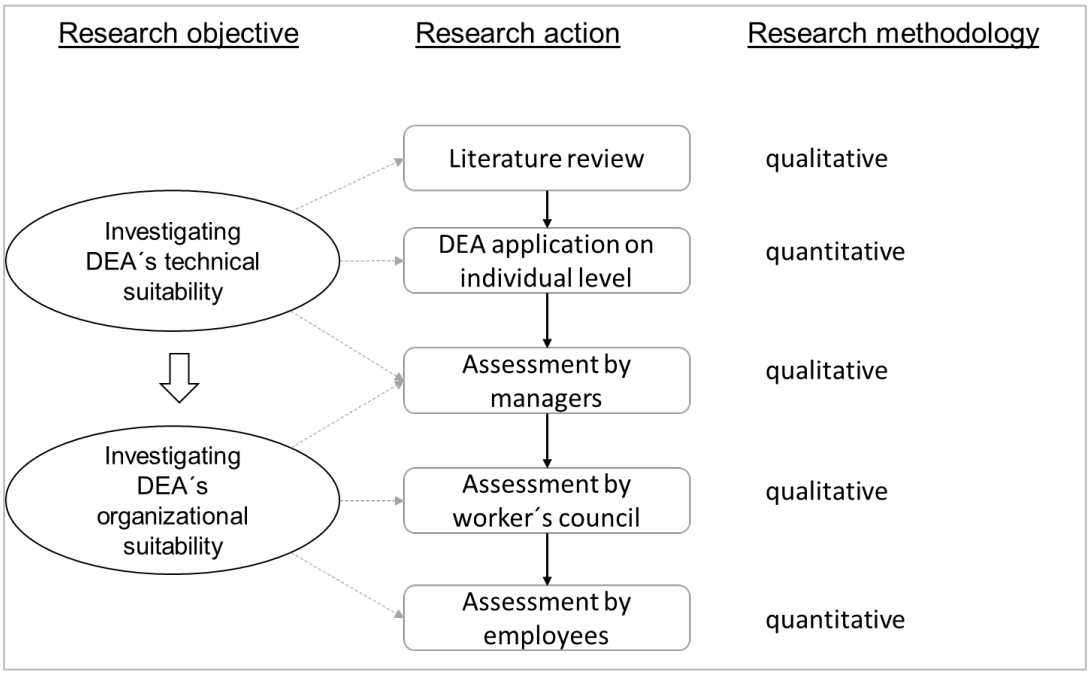


Figure 28: Sequential mixed-methods methodology

6.2.2.4 Research strategy: The case study

To achieve a reasonable level of coherence throughout the research design, the research strategy should link the research philosophy, the research methodology and the subsequent choice of data collection methods. Research literature provides a variety of research strategies, which are not necessarily mutually exclusive. For this research project, an embedded single case study approach was chosen. Hence, this chapter provides a justification for adapting a case study approach and illustrates the design of the case study.

Choice of research strategy

Among the research strategies that are used most often are surveys, experiments, archival research or action research. Although they are not mutually exclusive, most of them are either of qualitative or quantitative nature. The case study, on the other hand, generally uses a mix of qualitative and quantitative methods (Saunders et al., 2012). Yin (2017) even suggests that a case study always should rely on multiple sources of evidence, including interviews, documents, archival records or survey and by this deliberately triangulate the evidence from these multiple sources to confirm and corroborate the findings. Case studies, however, may also be applied as part of a larger mixed-methods approach. Thus, a common preconception is that case studies are just the exploratory stage of some other research strategy (Yin, 2017) . During the last

decades, case study research has become popular as a main research strategy which is able to integrate other strategies (e.g. a survey study) (Saunders et al., 2012). Yin (2017) argues that confinement to a case study forces the methods being mixed into an integrated mode. Moreover, case study research comprises an all-encompassing mode of inquiry and therefore case studies are not limited to being a data collection tactic alone or even a design feature alone. Considering the pragmatic philosophy and the mixed-methods approach that guide this research, the case study approach proves to be a reasonable choice.

Defining the "case"

An important requirement for the use of a case study design is to carefully define the case. According to Yin´s definition (2017) a case study "consist of an in-depth inquiry into a complex and specific phenomenon ("the case") within its real-world context." Yin (2013) also points out that the case should not be investigated in isolation but should be investigated in interaction with its context. Thus, case study research is appropriate to help understanding a real-world case when the understanding requires to involve important contextual conditions. This also distinguishes case studies from other research strategies. Experiments, for instance, separate a phenomenon from its context, attending only to the phenomenon of interest. Survey research also often struggles to investigate a phenomenon´s context since the number of items in a questionnaire are usually limited.

Initially, case studies merely focused on an individual person as the "case". In current studies, the "case" also can be some event or entity other than a single person. Other case studies focus on decisions or interventions as cases. In general, the case derives from the research question (Yin, 2017). For this study, the "case" can be defined as the application of DEA for evaluating employee performance in a German Cooperative bank. One major finding from the literature review is that the adequateness of a performance evaluation method not only depends on the procedure itself, but mainly on how it is perceived by the stakeholders, in particular by the employees (Chartered Institute of Personnel and Development, 2016a). Therefore, it is essential to study the case of DEA in interaction with its practical context. Hence, the practical context of this study is established by the requirements of the organization, but also by individuals and groups within the organizational context.

Designing the case study

A critical issue to the design of the case study is to define the case considering two distinct dimensions (Saunders et al., 2012):

a) single or multiple case(s)

b) holistic or embedded case

Considering the number of cases that are to be investigated, multiple-case studies are often favored since they are more likely to be stronger than single-case studies. The rationale behind multiple case studies is to replicate findings across cases (Saunders et al., 2012). Although the multiple-case approach often is preferred, there are several circumstances that indicate the use of a single-case design. Yin (2017) list five rationales for selecting a single-case study over a multiple-case study design:

- the single-case represents the critical test of a significant theory,

- the case represents an extreme case or an unusual case,

- the case is a common case,

- the case is a revelatory case,

- is case includes a longitudinal study.

For the application of DEA to evaluate employee performance in a cooperative bank, two of the mentioned rationales are applicable. First, the case is a common case. Thus, the study aims to capture the working procedures and conditions of a typical service organization offering professional services. Thus, the procedures as well as the input and output factors of the service production are not special to the organization but rather common and consistent for German banks (and also partly transferable to other service organizations). Yet, one requirement to carry out a case study is that the researcher needs sufficient access to the data for the potential case, for instance to interview people, review documents or records (Yin, 2017). This requirement makes the case also a revelatory case, where the researcher has the opportunity to observe and analyze a phenomenon merely inaccessible to social science inquiry. This is true for this case, since – although the case and procedures are common – the investigation of employee performance requests a lot of sensible data (including data on income and education of employees). Although the data was provided as linked-anonymous data, the data was not easily accessible, and the provision of the data was partly due to a trustful relationship with the bank resulting from previous research.

To address some of the concerns considering the validity and reliability of single-case findings, the assessment of DEA´s organizational suitability by the stakeholders was expanded to groups of managers and workers´ councils from other regional cooperative banks. According to Yin (2017) it would be a mistake to consider other cases that are similar to the original single-case, as if they were embedded subunits. Rather they would be considered part of a multiple-case design. However, from the researcher´s point of view, the other groups of managers and workers´ council do not form other cases, since the case is the application of DEA, which was only carried out once. Rather, by including other groups (or sub-units) (from similar organizations) in the study, only the context was enhanced but not the case. One of the first single-case studies including units outside the direct organizational context of the case was carried out by Lipset et al. (1956). Accordingly, the present research is considered a single-case research in an enhanced organizational context.

Another issue considering the design of the case study is the decision on a holistic or an embedded case study referring to the units or sub-units of analysis within the case. If the investigation is concerned only with the organization as a whole the case study is of holistic nature. If a number of sub-units within the organization are examined (e.g. conducting interviews with HR-managers) the case is labelled an embedded case study ((Saunders et al., 2012). As becomes apparent from the research objectives and methodology, the present study relies on holistic data collection for studying the main case but then calls on mixed-method data collection and analysis techniques to collect data from the embedded sub-units of analysis. Hence, this research is conducted as an embedded single-case study.

6.2.3 Data Collection and Analysis

Based on the research objective and its subsequent design, the data that are to be collected need to be defined. Further, in order to define the logic linking of the data to the research questions, analysis techniques for interpreting the findings are to be selected for this research.

6.2.3.1 Collecting and Analyzing case study evidence

Research evidence can come from various sources. Most researches classify research evidence on a first level into primary data and secondary data. Primary data are kind of new data that is collected specifically for the intended research purpose, whereas secondary data consist of data that has already been collected for some other purpose.

Both kinds of data, however, may provide useful evidence for research (Saunders et al., 2012). Among typical types of secondary data are archival records including organization´s data bases, reports of committees´, interview transcripts or media accounts (Saunders et al., 2012). Whilst secondary data are already available and therefore require no specific collection method, primary data can be obtained by various collection methods. Thus, primary data collection includes people (surveys and interviews), observation of events, physical documents (products such as student portfolios), and assessments (tests) (Mertens, 2014). Again, those methods are usually classified into quantitative and qualitative approaches for data collection. Comprehensive overviews of quantitative and qualitative collection methods including the discussion of their respective strengths and weaknesses are provided by Yin (2017) and Mertens (2014). As outlined before, this research draws on several research actions. With regards to these specific requirements a mixed-methods approach was chosen. Thus, the data collection methods are expected to serve this focus.

The research into DEA´s technical suitability is mainly carried out by applying DEA to archival performance data and by subsequently performing a sensitivity and portfolio analysis. Subsequently, the results of the DEA application are assessed by the distinct groups of stakeholders. For the assessment by managers and workers´ council, focus group interviews are conducted and analysed. In order to combine the advantages of the qualitative and the quantitative approach, the focus group interviews are followed by a questionnaire survey assessing employees´ perceptions of the proposed method. This way, it is possible to identify problems, concerns or areas of lacking information during the focus group interviews that could be made subject in the subsequent questionnaire. The questionnaire findings are analysed using descriptive statistics. Further, to identify cause-effect relationships, the results further are analysed using explanatory statistics including multiple regression analysis and Chi square analysis.

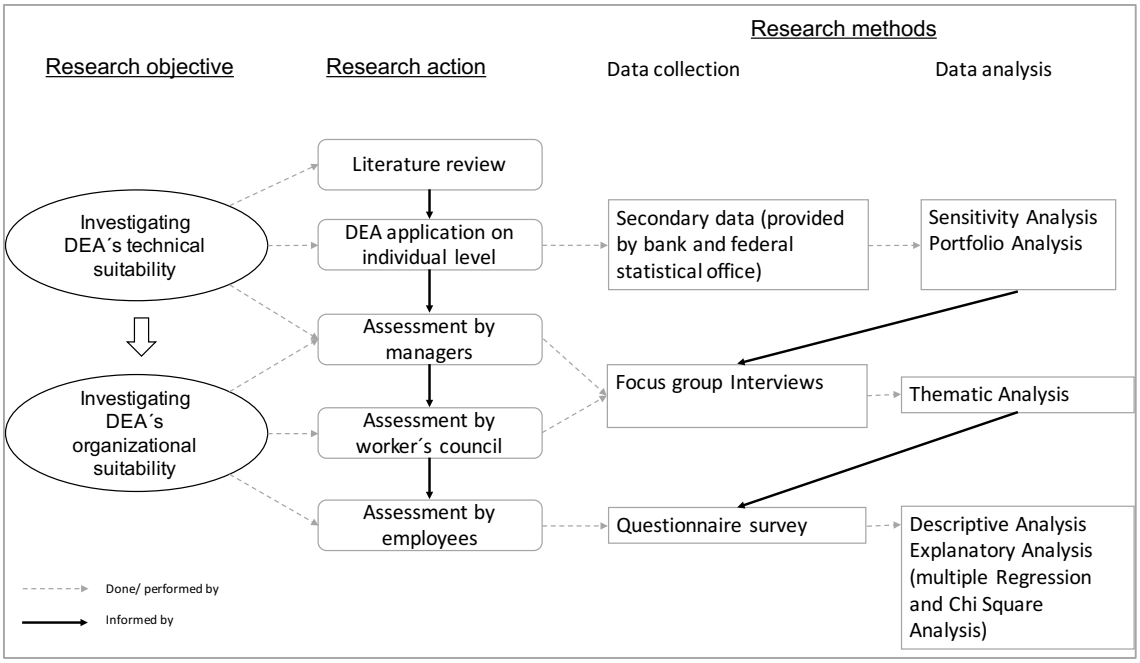


Figure 29: Methodological triangulation

The figure above shows the methodological triangulation that is used for data collection and analysis in this study. The methodological triangulation allows the researcher to better balance each method's strengths and weaknesses and eventually to increase research validity and reliability. The justification for choosing distinct methods is provided in the following sections.

6.2.3.2 DEA Application on for evaluating employee performance

Data collection

In a first step, DEA is applied in the organizational context of a German cooperative bank to evaluate employee performance. The performance data is mainly provided by the bank (except for data to capture environmental impacts (namely on the number of inhabitants in the catchment area), which was accessed through the federal statistical office). Hence, the application is based on secondary archival data, which is often used in research projects to base the collection of primary data on (Saunders et al., 2012). In this case, the advantage of using secondary data is the accessibility of data that generally are not easily available for researchers. Moreover, the provided data set consists of varying variables of personal and performance data that are of high quality considering detailedness and completeness (considering missing records) since its

primary use was aimed on the bank´s HR activities. One of the major drawbacks of using secondary data is the potential absence of key variables (Bryman, 2015). For this study, this is true for variables concerning environmental factors such as number of inhabitants or customer satisfaction. This issue was met by supplementing the bank´s data with other secondary data from the federal statistical office. Data on customer satisfaction (per employee), however, could not be obtained since this data is hardly ever collected on an individual level.

Data Analysis

After applying DEA to the performance data, the results of the application are analysed to assess DEA´s technical suitability. To perform DEA analysis, the software BANXIA Analyst is used. Since the results of the performance evaluation are of quantitative nature, the assessment whether DEA provides valid and reasonable results on a technical level should also be based on quantitative analysis (Paradi and Zhu, 2013). Therefore, a sensitivity analysis along with a portfolio analysis was performed. Considering the investigation of a statistical or mathematical model, sensitivity analysis is critical to gauge their relevance and plausibility. Ferretti et a. (2016) point out the importance of conducting a comprehensive sensitivity analysis by illustrating the case of Rogoff and Reinhert, two economists who published a paper claiming that GDP growth was slowed by high governmental debt. Other researchers later discovered that there was an error in an Excel spreadsheet they used to perform the analysis. Moreover, the choice of weights they used in their regression model was questioned as being unrealistic. Hence, the lack of a sensitivity analysis in economics can be a cause of significant problems. Despite its crucial importance to test the plausibility and technical suitability of a model, the sensitivity analysis is often overlooked or performed unsatisfactorily (Ferretti et al., 2016). This is also true for DEA research. Although various methods for testing the robustness of DEA´s results exist, there is no agreed procedure for sensitivity analysis in DEA (see chapter 2.4). Therefore, a comprehensive procedure to investigate the technical results of a DEA application on individual level is developed and applied. Further, a portfolio analysis is conducted to aggregate the individual results and to draw conclusions on organizational level. The application of DEA and the subsequent analysis of results is discussed throughout chapter 7.

6.2.3.3 Assessment by managers and workers´ council

Segmentation of stakeholders

To assess whether DEA is suitable to meet the organizational requirements to a performance evaluation that have been outlined in chapter 4, its perception by all stakeholders involved in the evaluation procedure is assessed. Thus, three distinct groups of stakeholders were identified:

- Management (management and line managers)
- Workers´ council
- Employees

The assessment focused on several criteria. First, stakeholders should evaluate whether the proposed DEA method is suitable to meet the administrative and developmental purposes that are defined by the organization. In this context, it is also assessed whether the approach is applicable for the purposes of determining performance targets and for calculating bonus payments. Second, the assessment focuses on the approach´s suitability to meet requirements that are generally associated with performance evaluation, including perceived fairness, objectivity and the provision of feedback information (see chapter 4). In this respect, it is also investigated whether the obtained results are comprehensible to all stakeholders. Third, a focus was placed on the method´s suitability to meet legal requirements. Since this topic required advanced knowledge concerning employment law, it should be particularly subject to the assessment with the workers´ council and HR managers. Finally, all stakeholders should evaluate whether the approach is perceived as superior, worse or even compared to other approaches or methods they know for evaluating employee performance. The table below provides an overview about the topics addressed.

	suitability concerning				comparison to other evaluation methods
	administrative purposes	developmental purposes	general requirements	legal requirements	
managers	xxx	xx	xx	xx	xxx
worker´s council	x	xx	xx	xxx	xxx
employees	x	xxx	xxx		xxx

xxx = high priority xx = medium priority x = low priority

Table 10: Priorities of survey for each group of stakeholders

According to their different background and intentions, the assessment sets different priorities for each group of stakeholders, which are marked in the table above. Considering the different priorities and purposes in each group and their distinct features (including group size, attitude or prior knowledge), it seems reasonable that the application of a single data collection method would not properly account for the distinct features of the groups under assessment and the slightly different purposes of the assessment.

Data collection

Since management and workers´ council form rather small groups (with less than ten members per group) a quantitative approach would most likely not provide valid results (Mertens, 2014). Moreover, the assessment by those distinct groups aims at investigating complex issues like the approach´s transferability to other job profiles that require rather open questions. Thus, a qualitative approach to data collection and analysis seems an appropriate choice for both groups. The most common qualitative methods for this purpose are participant observation, in-depth interviews and focus group interviews (Weathington et al., 2010). Since the assessment is focused on obtaining data on perceptions rather than on actions, observations of any kind would not provide the needed information. Hence, in-depth interviews and focus group interviews are worth considering. In-depth, semi-structured interviews are most advantageous when a large number of questions is to be answered, the questions are open and of rather complex nature and when the logic of questioning may be varied (Saunders et al., 2012). They may provide even more information than focus groups since not all participants may actively take part in a group discussion. In addition, certain personalities of participants in focus groups may influence the opinion of others or some members of the group feel to intimidate to speak. The last issue accounts in particular if topics are perceived as too personal (Krueger and Casey, 2009). Focus groups, on the other hand use group interaction as part of the method. Thus, participants are encouraged to talk to each other instead of the researcher asking each participant to respond to a question in turn. This makes the methodology particularly helpful to explore people´s knowledge and experience. More precisely, the group interaction may bring to light information that would be less easily accessible in a one to one interview and may take research in new and often unexpected directions (for instance, raising underlying issues and concerns that might otherwise be missed using other research methods). Moreover, focus group interviews allow the researcher to develop a deeper understanding why participants feel the way they do. This is because focus group discussions allow participants to probe

each other´s reasons for holding a certain view (Bryman, 2015). Group interviews provide information more quickly and often require a smaller amount of time than in-depth interviews. This, in turn allows the researcher to increase the size of the sample. Hence, focus group interviews were chosen for the assessment by managers and members of the workers´ council.

Data Analysis

The collection of qualitative data is usually linked to generating a large cumbersome database including field notes, interview transcripts or documents. Thus, a researcher needs a sound strategy for analyzing qualitative data (Bryman, 2015). Several strategies for qualitative data analysis have emerged over time. There are two general approaches to qualitative data analysis: the inductive or the deductive approach (the nature of inductive and deductive approaches has been discussed before. See chapter 6.2.2.1). Among the most used specific approaches to inductive analysis are grounded theory, template analysis and narrative analysis. A common approach in deductive analyses is pattern matching. A comprehensive analysis of these approaches is provided by Saunders et al. (2012).

Among the most widely used frameworks for analyzing qualitative data is grounded theory (Bryman, 2015). It is a rather inductive approach to develop a “grounded” theory around a core category that emerges from the researcher´s data. Since it was first proposed by Glaser and Strauss (1967), the exact nature of the procedure, however, has been modified by several scholars since and therefore varies between sources (Saunders et al., 2012). Despite varying procedures, there are two central features of grounded theory: the development of theory (or concepts) out of data and the iterative nature of the approach, meaning that data collection and analysis proceed in tandem, repeatedly referring back to each other (Bryman, 2015).

The structured procedure of grounded theory holds several points of reference for analyzing the data from the focus group interviews of this research. This accounts in particular for the process of coding and for theoretical saturation concerning the collection of data (see Bryman (2015). The process of data collection for this research, however, is of rather straightforward than of iterative nature. Thus, the features of the proposed DEA method are assessed among groups of managers and workers´ council based on a rather structured topic guide list. Since the assessment refers to predetermined features drawn from literature analysis (e.g. the method´s suitability to meet developmental purposes) data collection and data analysis of the interviews are performed in a sequential manner not referring back to each during the process of

collection. This is done because the purpose of the interviews mainly is to provide comparative generalizations (deductive) rather than building new theory (inductive). Thus, the researcher brings her research questions or a series of topics that derive from the research questions to the data. The obtained data, however, are also used inductively for identifying topics to inform the development of the subsequent questionnaire survey among employees.

Since the data collection procedure is not performed in an iterative manner and combines inductive and deductive elements, the rigorous procedure of grounded theory is not applied for data analysis. Although Bryman (2015) and Mertens (2014) point out that approaches to analyze qualitative data do not necessarily need to be linked to a specific approach. This is particularly true for research that is based on a pragmatic philosophy. On the other hand, Braun et al. (2012) argue, that a method that provides guidance and structure should support the analysis. Therefore, thematic analysis (TA), which was first proposed by Braun and Clarke in 2006, was used for the analysis of the interview data for this research including the following steps (Braun et al., 2012):

1) Getting familiar with the data
2) Generating initial codes
3) Discovering themes
4) Reviewing themes
5) Defining and naming themes
6) Writing the report

The preparation of the focus group interviews including the sampling process, the analysis of results following the generic approach by Braun and Clarke (2006) and the presentation and discussion of results are presented in chapter 8.1.

6.2.3.4 Assessment by employees

Data collection

Although, in DEA research there are already some studies concerned with measuring employee performance, there is no reported study about the employee´s perception of the methodology (see chapter 2.6.2). At the same time, employees´ perceptions are crucial for a performance evaluation to be accepted (Chartered Institute of Personnel and Development, 2016a). To address this lack of information and to investigate the

relationship of employees´ expectations and subsequent perceptions of a performance evaluation system (PES), the employee´s responses should be subject to a comprehensive empirical analysis enabling both descriptive and explanatory analysis. Thus, a survey as a prominent qualitative data collection approach seems to be an appropriate choice to assess employee´s perception of the proposed PES.

In general, a structured (or standardized) interview and a questionnaire are the two main ways to administer a survey. In a structured interview the questions are asked by an interviewer in a predetermined order. This brings the advantage of being able to ask more open questions and that questions may be explained to the respondents (Bryman, 2015). Moreover, response rates in personal interviews tend to be higher, in particular if questionnaires are rather long (Mertens, 2014). In questionnaire surveys usually no interviewer is present to ask questions. Thus, questions should rather be easy to answer. However, being asked questions by an interviewer may yield several problems. First, there is evidence in research that interviewer´s attributes including gender, race or social status can have an impact on respondent´s replies. Another disadvantage, that may be especially true for this research, is the social desirability effect. This effect refers to the issue that some respondents may provide answers that are related to their perception of the social desirability of those answers (Bryman, 2015). In this case, it is conceivable that respondents might provide answers of which they think they are desired by management or even society considering an innovative approach to performance measurement. This also leads to another problem of structured interviews, which lies in the fact that anonymity cannot fully be granted, which is at least the case in face to face interviews.

Considering the drawbacks of structured interviews, the method of a self-completion questionnaire is selected for data collection among employees. Besides accounting for the shortcomings of a structured interview (in reducing interviewer bias and the risk of receiving social desirable answers) the questionnaire brings the advantages of being cheaper and quicker to administer (Bryman, 2015). However, to obtain sufficient information to perform a subsequent statistical analysis, some of the traditional shortcomings of questionnaire studies need to be addressed. One major issue in questionnaire studies is, that questions usually cannot be explained due to the absence of an interviewer. Since the DEA application and its subsequent results are rather complex, they have to be explained to the respondents in advance of the survey anyway along with the purpose of the study (Saunders et al., 2012). Thus, it is expedient to have the respondents who agree to participate complete the survey on-site. In doing so, two other problems can be addressed. Response rates, which tend to decrease when using

delivery methods like letters or e-mail (Mertens, 2014), can be held quite high. Further, the researcher is present during the completion of the questionnaire. Thus, she can explain the meaning of questions if respondents had trouble to answer them. Considering the advantages of an on-site completion and the presence of the researcher for collection, the risk of reinforcing interviewer bias to a certain degree is acknowledged and accepted by the researcher.

Data Analysis

The assessment by employees is of descriptive nature for one part. Thus, the participants of the questionnaire study are requested to evaluate the proposed method considering its suitability to meet general requirements and to account for administrative and developmental purposes. For the analysis of this part of the assessment, descriptive statistics are performed presenting the proportions of occurrences in frequency tables. The second part of the assessment seeks to identify cause-effect relationships that link the requirements of a performance evaluation to its actual perception by the employees. Since this part of the investigation is rather of explanatory nature, the analysis was performed using multiple regression analysis and chi square analysis. The statistical analysis is carried out using the software SPSS. The questionnaire design, the sampling and the subsequent descriptive and explanatory analysis are presented in chapter 8.2.

6.2.4 Ethical Considerations

The research design, in particular data collection and analysis, often are associated with a range of ethical issues (Saunders et al., 2012). As cited in Bryman (2015) Diener and Crandall (1978) list four main areas where issues considering research ethics in social science typically arise:

- Harm to participants
- Lack of informed consent
- Invasion of privacy
- Deception

One major principle for avoiding doing harm to participants is to maintain identities and records of individuals confidential. For this research, this applies to the provision of performance data by the participating bank and to the participants taking part in the assessment of DEA´s results. The performance data that was provided to the researcher

was archival data and was provided as linked-anonymous data. Therefore, no conclusions about personal identities or current performances of employees could be drawn by the researcher. This procedure also ensured that the data could not be used by the bank to support measures or decisions with respect to particular individuals. The provided data included personal data (e.g. work-experience, salary, education) but no sensitive data as defined in the Data Protection Act 1998.

The assessment of DEA´s organizational suitability was carried out by focus group interviews among managers and workers´ council and by a questionnaire survey among employees. The participation in the assessment was voluntary. In a participant information sheet the participants were informed about the voluntary nature of the participation and that they could withdraw from the evaluation at any point. The focus group discussions were not recorded upon the wish of the bank´s management. Rather, the discussions were logged and transcribed afterwards. The names of the participants were coded so that no personal names were used in the notes and transcripts. The transcripts were provided to the participants afterwards. Also, the employees taking part in the questionnaire survey were provided with the participant information sheet. The survey was completed anonymously. Each questionnaire was coded for further analysis. For the analysis and results discussion, the individual answers were summarized and only the descriptive summaries are displayed in the thesis. Thus, it is not possible to draw conclusions on individuals from the results.

The participant information sheet was handed out to all participants along with a form on informed consent. The informed consent sheet informed the participants on the purpose of the research, the anonymity and voluntary nature of their participation and their possibility to withdraw at any time. To ensure informed consent, the participants signed the form in advance of the assessment. By providing informed consent, the participants, however, do not lose their right to privacy. To prevent transgressions of privacy, the participants of the questionnaire survey had the opportunity to select “no answer” for each questions. The participants of the focus group interviews of course had the possibility to not answer the question or engage in the discussion if they felt the question was too private.

Deception in research generally occurs if the researcher presents her work as something other than what it is really is. To various degrees, deception is often applied in social psychology experiments but is not merely exclusive for this are of research (Bryman, 2015). Again, to prevent deception, all participants of this research were informed of its nature and its purpose by the participant information sheet. This research project further adheres with the Anglia Ruskin University´s code of research practice. The research

project including the participant information sheet and the informed consent form were presented to and approved by the University´s Research Committee.

6.2.5 Validity and Reliability

The research design is supposed to represent a logical set of research actions, that are interrelated within a research process. They also constitute essential criteria for the successful evaluation of the accuracy and rigor of the research (Saunders et al., 2012). Hence, the research design including its methods for data collection and analysis should be valid, in terms of producing accurate findings which clearly represent what they are about to produce. The obtained data should also be reliable, meaning that results of a study should be repeatable, for instance by a follow-up analysis or by studies from other researchers (Bryman, 2015). To establish the quality of empirical social research, Yin (2017) proposes four criteria including construct validity, internal and external validity and reliability.

Construct validity is mostly associated with the question whether a measure that is devised to a concept does adequately reflect the concept it is supposed to measure. This criterion particularly accounts for quantitative measures (Bryman, 2015). Yin (2017) notes that case study research is often criticised for failing to develop a sufficiently operational set of measures and for using “subjective” judgments for collect the data. Therefore, to increase construct validity in case studies, he suggests to use of multiple sources of evidence, in a manner encouraging convergent lines of inquiry and to establish a chain of evidence, also relevant during data collection. To account for this suggestions, this research is based on a mixed methods approach for data collection and analysis. This methodological triangulation is used to facilitate the validation of the obtained data through cross verification.

Internal validity mainly relates to the issue of causality and the question whether a conclusion that incorporates a causal relationship is robust. Hence, internal validity mostly concerns explanatory studies (Bryman, 2015). For this research, the question of internal validity is mostly applicable for the explanatory part of the questionnaire survey. To ensure internal validity, the significance of the causal relationships was tested determining a confidence level of alpha = .05. External validity, on the other hand, refers to the issue whether the results of the study are generalizable beyond the specific research context (Bryman, 2015). Especially case studies have been accused of their lack of rigor and little basis for scientific generalization (Yin, 2017). Thus, the researcher is aware of the fact that it should be avoided to draw statistical generalizations from a

case study to a population. Rather, Yin (2017) claims that a case study gives the opportunity to shed empirical light on some theoretical concepts or principles. Thus, he labels the comparison of the results of a case study to a previously developed theory as "analytic generalisation". Therefore, although no statistical generalizations are drawn, this research may contribute to make some analytic generalizations by comparing the findings from the case study to the previously developed theory and subsequent research questions. To enhance external validity, the organizational context of the case study, meaning the assessment of the proposed method by the stakeholders, was expanded by involving additional participants (e.g. managers and workers´ councils from other regional cooperative banks) into the assessment.

The reliability of research is mostly concerned with the repeatability of a study (Bryman, 2015). A reliable study should therefore ensure that, if a later researcher follows the same procedures as described by an earlier researcher, she will arrive at the same findings and conclusions (Yin, 2017). However, findings from qualitative research including in-depth interviews or focus group discussions are not necessarily intended to be reproducible since they reflect reality at a time they were collected under circumstances that may be subject to change (Saunders et al., 2012). Also, case studies are hardly repeatable under the same conditions (Yin, 2017). Despite the impossibility to exactly reproduce qualitative findings the researcher should still position her work to reflect a concern over reliability. Thus, for qualitative research it is suggested to provide a detailed account of the procedures, the methods and decision points of the study. To account for this requirement, the chapter at hand and the chapters illustrating the analysis process present, discuss and document these procedures in detail. Further, all interviews were documented and protocolled. The interview protocols were also handed out to the participants. In addition, the topics of the qualitative assessment were documented in a topic guide list. For the quantitative part of this research, namely the questionnaire study, reliability is mostly concerned with the robustness of the questionnaire and whether it will produce consistent findings (Saunders et al., 2012). Saunders et al. (2012) list several approaches to assess reliability in questionnaires, among them the calculation of Cronbach´s alpha for internal consistency. This method was applied for the questionnaire used in this study. The test produced alpha values ranging from .726 to .940 for all sections concerned with the assessment of the proposed approach (see chapter 8.3.2.1). The limit of .7 is generally accepted as a lower bound for consistency. Thus, the questionnaire proves to be reliable by the means of Cronbach's alpha.

6.2.6 Methodological Limitations of the research project

In the previous chapters, the underlying philosophy of this research project has been presented along with the associated research design including the research strategy and the methods for data collection and analysis. The research methodology was chosen regarding its adequateness to address the research questions and has already proven its usefulness in several other studies. However, the choice of methodology imposes some major limitations to this research, which will be discussed in the following .

One major limitation lies with the pragmatic research philosophy. Whilst the justification to base this research on a pragmatic philosophy has been presented before, pragmatism also has some shortcomings. Thus, pragmatic research is hardly known for promoting fundamental or revolutionary change in society rather than incremental change. Further, due to its rather practical nature, it often fails to provide solutions to philosophical disputes (Johnson and Onwuegbuzie, 2004). Hence, the findings of this research aim at contributing to find a rather applicable solution to evaluate employee performance adequately. In doing so, the results of this research may provide a basis for modifying current evaluation procedures but may not bring a revolutionary change in evaluating employee performance. Moreover, the findings may inform the design and development of HR practices but may not impinge on current philosophical debates including debates on the nature of services (see 5.2) or the role of HRM (see 3.1.3).

A mixed methods approach, which was applied for data collection and analysis, is often associated with pragmatist research overcoming the qualitative/ quantitative divide. The use of both, qualitative and quantitative methods therefore places high demands on the researcher considering a good understanding of both methods and requires the researcher to adopt different roles (Feilzer, 2010). Moreover, mixed-methods research generally takes more time for data collection and analysis and therefore generally focuses on a smaller sample (Johnson and Onwuegbuzie, 2004). Also, it may be problematic to integrate different methods into consistent findings (Feilzer, 2010). The applied methods for data collection and analysis have been presented in the methods at hand and in chapters 7 and 8 to demonstrate a profound understanding of the respective methods. Further, the data collection process was strictly oriented towards addressing the research questions that were derived from literature. Data analysis and the subsequent integration of findings followed the principle of triangulation, and were aimed at answering the research questions. Thus, research objectives and research questions served as orientation throughout the research process and supported the integration of findings from different sources of evidence. However, the process of data collection (in particular carrying out the focus group interviews and acquiring participants

for the questionnaire survey) was time-consuming. Thus, the sample size may have been larger if a mono-method approach would have been chosen.

Another major limitation relates to the generalizability of findings from this study. As outlined before, the case study approach often is associated with limited generalization to the population (Yin, 2013). To strengthen generalizability and to boost confidence in the findings, Yin (2013) suggests a qualitative comparative analysis that involves a systematic comparison of within-case configurations or sets of interventions. To allow for this kind of analysis, the organizational context of the case was expanded by involving other groups of managers and workers´ councils than form the particular bank that provided the data into the assessment. Moreover, both Yin (2017) and Bryman (2015) point out that case study research may not be generalizable to populations but rather allows to draw analytic generalizations to inform theory and practices. However, it should be noted that the findings from this study, in particular the perception of the proposed method by managers, workers´ councils and employees are not generalizable to the population of Germany or to all service organizations.

Finally, all empirical research is subject to the quality of the empirical data. In this research project, data collection relied on three major sources of evidence: performance data provided by the bank under investigation, qualitative data from the focus group interviews and quantitative data from the questionnaire survey. Considering the performance data, there is a limitation concerning the number of performance criteria. Thus, the bank only provided data they had readily available (environmental data was additionally obtained from the federal statistical office). In consequence, the investigated performance process was restricted to this data, meaning that some performance indicators that would have also been worth investigating (e.g. customer satisfaction per employee) could not be included. Considering the focus group interviews, a major limitation lies with the facilitation of the transcription and recording of the interviews. For recording interviews, the researcher needs to seek permission from the participants (Saunders et al., 2012). Several groups (both bank´s managers and workers´ councils) did not permit voice recording, therefore, to ensure a consistent procedure, all interviews were documented by taking notes. The notes were transcribed documented in an interview protocol immediately afterwards and the transcripts protocols were provided to the participants for verification. Since it would not be possible to take notes and to moderate a focus group discussion, the researcher took part in the focus groups as mere observer (see 8.2.1). By relying exclusively on notes, there is a chance that some of what was being said or other details including voice tones or phrases were lost. In addition, a line-by-line transcription was not possible. Regarding the questionnaire study,

to ensure a high response rate, the questionnaire was limited to 50 questions addressing topics and characteristics that were identified as important by literature research and the previous Focus Group Interviews. However, more questions on the topics addressed or even the inclusion of other topics could have altered the results.

Further, as a general shortcoming of qualitative interviews, the participants may not have provided all relevant information or may have been reluctant to put forward an opinion that they think of as “unwelcome”. As in all quantitative research, the questionnaire survey may be limited to the characteristics and size of the sample. Moreover, the topic of the survey was rather complex and therefore there is the risk that not all participants understood the rationale and results of the proposed method. To address this risk, the questionnaire was completed on-site with the researcher present to ask questions. However, it may not be ruled out completely that issues concerning the complexity of the method under investigation occurred among respondents.

7 Applying DEA to evaluate employee performance: a case study

After having discussed the theoretical foundations of DEA, Performance Management and the nature of Banking Services, this section illustrates the application of DEA for assessing employee performance in a German Cooperative bank. Since the performance assessment is an integral part of a performance evaluation system (PES), the procedure is closely oriented towards the model developed in 4.2. The application was carried out as a case study.

The section´s main focus is on the application of the DEA methodology. In this respect, the background and the purpose of the investigation will be illustrated. After that, the application of the DEA methodology is described and the results are analysed. Taking into account the outcome of a subsequent sensitivity analysis, the results are reviewed in detail. Based on this review, recommendations for HRM are derived and discussed on individual and organizational level. Finally, conclusions for implementing a DEA-based PES are drawn from the case study.

7.1 Performance Planning

To investigate DEA´s technical and organizational suitability for evaluating employee performance in the service industry, the methodology was applied to a German Cooperative bank. It should be noted that the name of the bank is not published due to confidential reasons. The investigation was carried out as a case study. In literature, researchers recommend the use of the case study method, when research seeks a deeper understanding of key issues or methods in a specific context (Haynes and Fryer, 2000). Since the aim of this research is to investigate the suitability of DEA in the context of measuring employee performance a case study approach seems an appropriate method to be applied (see section 1.5. for a detailed specification). In the following sections, background information on the analysed Cooperative bank will be outlined. Subsequently, the purpose and the focus of the performance evaluation, which were determined in cooperation with the bank´s management, will be examined more closely. Furthermore, the definition of the service process, in order to derive performance standards and performance measures, will be illustrated.

7.1.1 Background of the case study

Compared to an average German Cooperative bank, the analysed Cooperative bank is a rather small sized institution located in the northeast region of Germany. In 2013, when the study started, the bank then had about 100 employees, more than 8,000 customers and 15 branches. The branches were situated in small and medium town areas (about 10,000 to 35,000 inhabitants) but also in small village areas (about 500 to 2,000 inhabitants). The bank´s balance sheet was about 300 million EURO. Due to a strong regional orientation, the bank´s strategic focus was in particular on retail and SME business. Considering the distinct types of banking services, the bank offered all three archetypes of services along Silvestro´s (1999) service process.

Although the institution managed a small growth during the financial crisis and its aftermath, it was facing major strategic challenges in the near future. One of them was a demographic challenge, since the average age of customers has risen significantly during the last decade. In this respect, they were also facing a competitive challenge caused by direct banks, which particularly attract young customers. Direct banks do not have an own branch network, but provide their services solely in direct communication with customers, mainly by online-services. Although their services are mainly based on “mass services” and “service factory” services, they have become serious competitors for traditional institutions like Cooperative banks with a large branch network. To meet those challenges, the bank´s management considered professional services and service shops as their main competitive advantage over direct banks.

Hence, the bank´s management pursued two directions of thrust. To emphasize their regional focus and point out their competitive advantage (for instance, offering tailored professional services), they needed to preserve their local branch network and improve direct communication with the customers. Thus, employees were encouraged to enhance the number and quality of meetings with customers. On the other hand, the cost structure of the branch network needed to be improved in order to maintain the network. In this respect, efficiency considerations had not only to be made on branch level, but also on employee level. Thus, the bank´s management recognized a need to decompose their strategy down to employee level and to link it to employee´s performance. Since the strategic targets were not necessarily complementary, the bank´s management had to determine which direction of thrust to emphasize in case of conflicting targets. Therefore, a hierarchy of targets was developed:

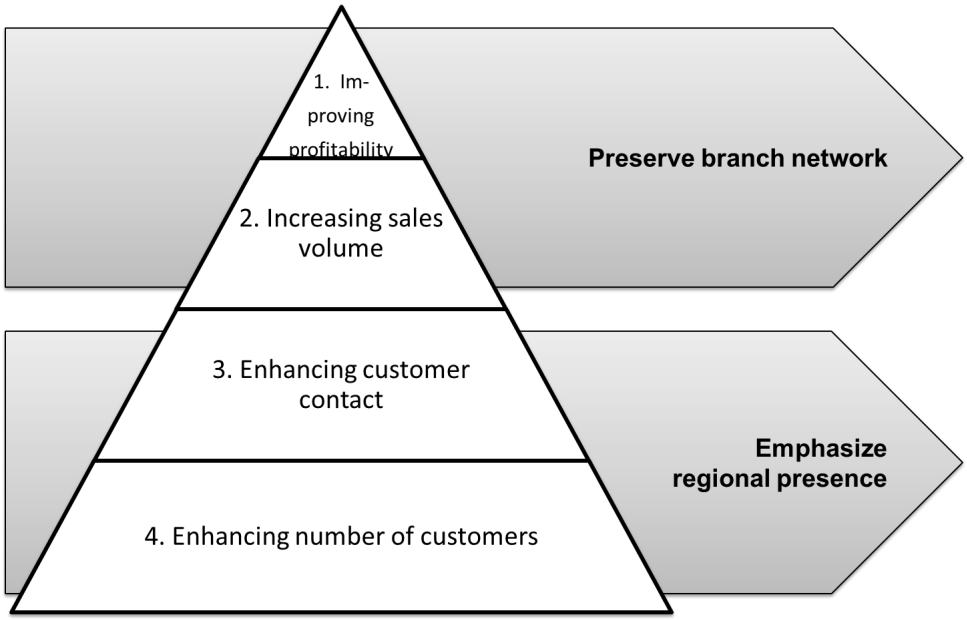


Figure 30: Hierarchy of strategic targets

Since management felt that it is of major importance to preserve the branch network in order to maintain their competitive advantage, they made profitability top priority followed by increasing sales volume. Adapting the rationale of state-of-the-art in service sector research (see 5.1), profitability and perceived quality should not be considered separately, since a changed set of inputs or outputs may alter the customer´s perception of the service. Thus, it would have been appropriate to include a target related to customer satisfaction. In literature, it is widely agreed that the use of customer satisfaction as a measure is particularly appropriate in the service sector where organizational dynamics have a direct impact on the customers (Schneider, 1994). However, an important precondition is that the data are collected systematically and longitudinally (Guest, 1997; Haynes and Fryer, 2000). Until the start of the case study, the bank did not collect suitable data on customer satisfaction neither on a regular basis nor on employee level, which would have made the formulation of a related target obsolete. To emphasize their regional presence, personal contact to existing customers should be enhanced as number three priority. In addition, new customers shall be acquired to address the demographic challenge.

7.1.2 Defining the purpose and requirements of the performance evaluation

In coherence with the procedure that is outlined in 4.3, the first step to develop a PES is to define its purpose and derive the subsequent requirements. For this purpose, several meetings were held with the bank´s management. In general, the bank´s management was interested in a PES that enables the assessment of employee´s performance on a continuous basis. The bank´s employees are initially classified into front office and back office employees. Front office employees again are classified into service employees (working almost exclusively at the counter) and account managers, whose tasks include the full-scale customer consulting and support. Whilst service employees´ tasks are mainly focused on providing mass services, account managers´ focus is on offering professional services or "service shop" services. Since professional services and "service shop" services require an advanced approach for evaluating employee performance, it was decided that the DEA approach should be applied to the group of account managers for the case study. Once the PES has been tested and evaluated, it should be implemented for all groups of employees.

Regarding the strategic directions of thrust, the PES should serve both administrative and developmental purposes. For administrative purposes, the results of PES should support decisions about promotions and about employee´s transfers to other branches. In this respect, the performance evaluation should in particular give evidence whether an employee has performed to her full potential or should be assigned to another branch (e.g. with a larger territory). Additionally, the results should be used to determine bonuses, which are paid as a variable pay component once a year.

For developmental purposes, the performance evaluation should identify areas of improvement and training needs. Furthermore, areas of "best practice" should be identified, pointing out where an employee has excelled in performance and may serve as a peer to others. The results should enable an individual counselling identifying individual strengths and weaknesses and drafting a personal development plan. To ensure a holistic consideration, the evaluation should take several performance standards into account. Thus, it was decided to consider multiple performance dimensions. Additionally, the results should be used to determine individual performance targets. The bank has already been using performance targets for personnel development and controlling, which were previously determined by an overall decision or "gut"-decisions.

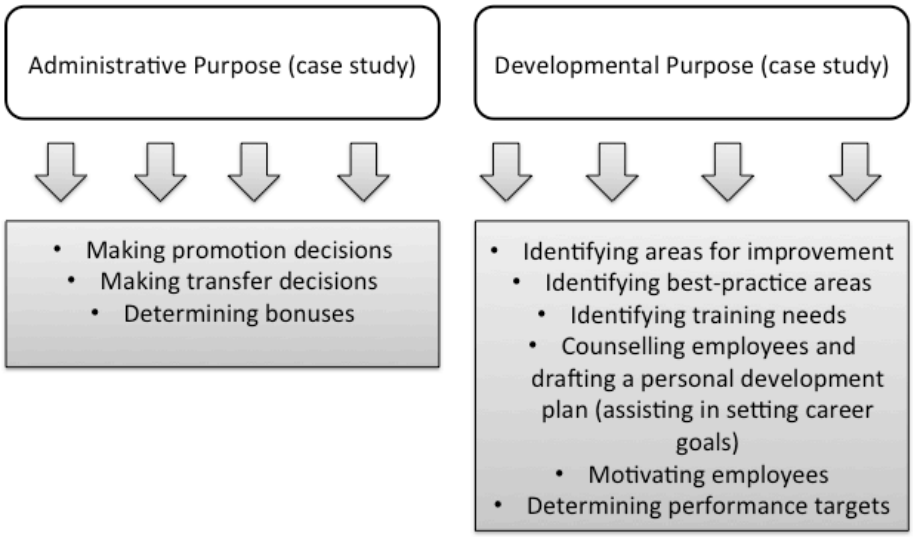


Figure 31: purposes of the PES - case study

Subsequently, the PES needs to meet the requirements outlined in 4. A general requirement to performance evaluation is perceived fairness. This is especially important, since the bank´s management reported employees´ dissatisfaction with the traditional process of determining performance targets. As there have been few objective standards yet, employees perceived this process as rather unfair. Supervisors also reported a growing dissatisfaction with the traditional procedure, as they lacked objective guidelines to determine performance targets. An additional requirement was the usage of balanced performance measures. Due to the complexity of banking services, this concerns in particular the usage of multiple input and output criteria. To enable the provision of individual feedback information and the identification of areas of improvement and best-practice, a comparative "to-best" approach was favoured by management. In order to determine bonuses, the results of the PES should be linked to the compensation process.

7.1.3 Defining the performance process and deducting performance dimensions

Although the application of DEA was - for the case study - initially limited to the group of account managers, the performance process of professional banking services still is increasingly complex. In order to grasp the complex nature of this process and to obtain results that are acceptable to both, managers and employees, the overall performance process was segmented into several sub-processes (performance dimensions). Thus, performance can be investigated from an individual and an overall perspective. A rather psychological advantage of taking multiple performance dimensions into account is that

employees may be more likely to accept suggestions on how to improve in some performance dimensions when they showed good performance in others (Paradi et al., 2011).

A top-down approach was chosen to deduct performance dimensions from the overall performance process. It was oriented towards the approach of Bradford et al. (1969). By this, the whole performance process is divided into several “transformation stages”, whereby the word “transformation” refers to the transformation of inputs into outputs (Westermann and Johnson, 1999). Thus, five dimensions, which are referred to as “transformations stages (TS)” in the following, were identified:

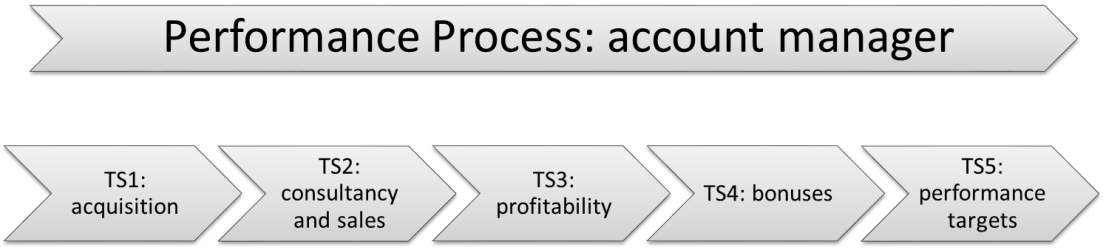


Figure 32: performance process - case study

Transformation stage 1 (“TS1”) deals with the employees’ ability to acquire new customers or to arrange appointments with existing customers. Transformation stage 2 investigates how successful the employees seize these appointments to counsel or to sell financial products. To capture the interrelation between the two stages, the outputs of TS1 serve as inputs for TS2 and with this are intermediate factors. Transformation stage 3 is a rather financial perspective and describes the relation of total costs to total revenues. In other words, this stage displays the employees` profitability by using the measures that were suggested by Grönroos and Ojasalo (2004) to capture the profitability of services. TS3 explicitly addresses an economical point of view and refers to the management´s strategic number one target to improve the branch network´s cost structure. This seems reasonable, for it is conceivable that an employee is very successful in acquiring new customers and in counseling them, but achieves those results by producing an excessively high amount of total costs. TS4 focuses on the determination of bonuses. So far, bonuses were determined by line managers on a rather subjective basis. According to management and workers´ council, bonus decisions have been subject to discussions and complaints ever since, as employees did not trust the process to be fair. Thus, bonuses should be calculated in a more objective manner taking into account the bank´s financial investment in relation to the employees´ contribution to

the bank´s added value. TS5 deals with the determination of individual performance targets. The use of performance targets for administrative and developmental purposes is already a lengthy tradition within in the bank´s HR practices. As in the case of bonus determination, the process of defining performance targets has so far been a rather subjective one. To enable decisions on a more objective basis, performance targets should be determined taking into account the employees´ individual working conditions (number and nature of customers for example) and their working experience in relation to their contribution to the bank´s revenues.

7.1.4 Defining performance measures

To define performance measures for each performance dimension, a measurement framework oriented towards the framework provided by Fitzgerald et al. (1991) was developed. It is based on the premise that there are two distinct types of performance variables: those that focus on results and those that focus on determinants (see 3.2.6). The variables that focus on results are subsequently referred to as output variables, the variables that focus on determinants as input variables. Thus, the rather process-based approach of Bradford et al. (1969) was combined with the rather hierarchical approach of Fitzgerald et al. (1969).

The performance measures were defined based on the guidelines provided by Thor (1993) and Neely et al. (1997). Thus, the following guiding principles applied:

- The measures are linked to strategy/ directions of thrust
- The measures enable the provision of timely and accurate feedback.
- The measures reflect the performance process
- The measures are objective
- The measures are easy to collect (uses data already available)
- The measures do not stimulate undesirable behaviour
- The measures are relevant to all employees under investigation

Considering these guidelines, several performance measures were defined for each transformation stage.

TS1: acquisition

To evaluate employee´s ability to acquire new customers or to make appointments with existing customers, several measures were taken into account. The results of the employee´s efforts to acquire customers are twofold. On the one hand, the number of appointments is an important measure. On the other hand, the success of those appointments should also be evaluated. Thus, the number of products sold by each employee was included. Considering the determinants of this stage, three variables seemed reasonable: To capture the potential of existing customers, the total number of customers served per employee is a crucial measure. Additionally, to capture the potential of new customers, the number of inhabitants within the branch´s catchment area was included. Although this is a measure the employee cannot control, it may influence the employee´s success in acquiring new customers. Another input measure that may have a significant influence on the output is the total volume of a customer´s deposit, since customers with a rather high volume of deposits may be more likely to buy financial products.

TS1: acquisition			
	Title	**Abbreviation**	**What is being measured**
Inputs	**Customers**	CUST	employee´s total number of customers served
	Inhabitants per employee	INHAB	Number of inhabitants living within the branch´s catchment area
	Active volume of customer´s deposits	VOLACT	Sum of the total active volume of deposits of the employee´s customers
Outputs	**Appointments**	APP	Number of appointments for financial consultations per year
	Selling	SELL	Number of financial products sold per year

Table 11: performance measures TS1

TS2: Consulting and sales

The determinants of the employee´s performance concerning consulting customers and selling financial products are merely the results of the previous stage. Thus, the numbers of appointments and financial products sold were defined as input measures. Two measures that obviously have an impact on the employee´s success in consultancy and sales are the total volume of customer's deposits and the total liability of customer's deposits that are handled by the employee. Those measures can be interpreted both ways: as input or as output measures. On the one hand, it may be easier to sell a high

volume of financial products if an employee handles a rather large volume of deposits or liabilities. Subsequently, the two measures should be treated as an input. On the other hand, the total volume of deposits and liabilities may be a direct result of the employee´s ability to consult and sell financial products. In consequence, the measures should be treated as an output. After an intense discussion about those two possibilities, the banks management decided to treat both measures as outputs. In addition, three other output measures were defined, which represent the bank´s major financial products: liabilities resulting from building societies, insurances and investments. A rather long-term output resulting from this stage is customer satisfaction. The inclusion of customer´s satisfaction would not only acknowledge its strategic importance, but may also have a behavioural influence (especially when it comes to the sale of high-risk products). The management strongly supported the inclusion of this variable. Unfortunately, the bank lacked a sufficient data base for this purpose. Thus, data on customer satisfaction were more than three years old. Additionally, the data have not been collected on individual level, but on branch level. Since this variable is of crucial importance, it was agreed to collect it on individual level on a regular basis in the future.

TS2: consultancy and sales			
	Title	**Abbreviation**	**What is being measured**
Inputs	**Appointments**	APP	Number of appointments for financial consultation
	Selling	SELL	Number of financial products
Outputs	**Active volume of customer´s deposits**	VOLACT	Sum of the total active volume of deposits of the employee´s customers
	Liability volume of customer´s deposits	VOLL	Sum of the total liability volume of deposits of the employee´s customers
	Volume (Liabilities) building society "Schwäbisch Hall"	BSH	Total volume (Liabilities) building society "Schwäbisch Hall" handled
	Volume (Liabilities) "RuV" Insurance	RuV	Total volume (Liabilities) "RuV" Insurance handled
	Volume (Liabilities) investment company "Union Investment"	Union	Total volume (Liabilities) investment company "Union Investment" handled

Table 12: performance measures TS2

TS3: profitability

To evaluate the relation of employee´s costs in relation to the revenue they are generating adequate cost variables had to be defined as inputs. In general, the labour costs result from salary expenses including variable remuneration and expenses for training. Variable remuneration is paid as bonus at the end of each fiscal year on an optional basis. On the other hand, the revenues account managers generate can be categorized into two groups: revenues that result from interest earnings and revenues that result from commissions. The first group is referred to as contribution margin II and the latter on as commission earnings.

TS3: profitability: Relation of employee´s total costs to total revenues achieved			
	Title	**Abbreviation**	**What is being measured**
Inputs	**Salary**	SAL	Salary
	Training	TRAI	Sum of expenditures for training
	Bonus	Bonus	Sum of bonus payments received
Outputs	**Contribution margin II**	CMII	Contribution margin II resulting from interest earnings
	Commission earnings	CommE	Sum of commissions earned

Table 13: performance measures TS3

TS4: Bonuses

To prevent a rather arbitrarily allocation of bonus payments based on “gut-feeling”, the performance evaluation shall also serve as a basis to calculate fair bonuses. As DEA provides a single score, this score will be the basis to calculate an employee´s bonus share. Like TS3 “profitability”, the determinants for bonus payments are the labour costs. Although training expenses may influence the outcome, management and workers´ council indicated that those expenses should not be used as an input as this may stimulate undesirable behaviour (e.g. employees refusing training offerings in favour of a rather high bonus payment). Again, contribution margin II and commission earnings are the output factors.

TS4: bonuses			
	Title	**Abbreviation**	**What is being measured**
Inputs	**Salary**	SAL	Salary
	Bonus	Bonus	Sum of bonus payments received
Outputs	**Contribution margin II**	CMII	Contribution margin II resulting from interest earnings
	Commission earnings	CommE	Sum of commissions earned

Table 14: performance measures TS 4

TS5: performance targets

To determine performance targets that relate to the employee´s individual internal and environmental conditions, the determinants were chosen in coherence to the variables that are used for TS1 “acquisition”. Thus, environmental conditions such as the number of inhabitants are taken into account as well as internal conditions such as the total volume of deposits handled by the employee. It was also agreed to include salary as an input factor, since this measure is able to address two issues. First, it reflects the employee´s experience (without having to include experience as a separate measure). Second, it emphasizes the organization´s philosophy that a higher salary is associated with a better performance. It was also suggested to include training expenses as an input measure. After a discussion this was finally rejected since most managers felt that including training expenses may contribute to stimulate undesirable behaviour (e.g. refusing training). Additionally, employee´s efforts and motivation to achieve performance targets should not be countered by using training as determinant.

TS5: performance targets			
	Title	**Abbreviation**	**What is being measured**
Inputs	**Customers per employee**	CUST	Total number of customers served per employee
	Inhabitants per employee	INHAB	Number of inhabitants living within the branch´s catchment area
	Salary	SAL	Salary
	Active volume of customer´s deposits	VOLACT	Sum of the total active volume of deposits of the employee´s customers
Outputs	**Contribution margin II**	CMII	Contribution margin II resulting from interest earnings
	Commission earnings	CommE	Sum of commissions earned

Table 15: performance measures TS5

The proposed stages and measures were subsequently discussed with the workers´ council, who eventually approved the whole proposal.

7.2 Performance Execution: Data collection and model building

To prepare data collection all measures were documented using a modified "performance measure record sheet" proposed by Neely et al. (Neely et al., 1997). The purpose of this documentation was to ensure that all implications of the proposed measures were considered. Besides that, the record sheet could also serve for communicating the measures to the employees. Two exemplary record sheets for the measure "inhabitants" and "total active volume" are displayed below.

Title	**Inhabitants per employee**
Abbreviation	INHAB
What is being measured?	Number of inhabitants living within the branch´s catchment area
Formula (if applicable)	n.a.
Frequency	annual
Source	Federal Statistical Office

Table 16: performance measure record sheet (measure "inhabitants")

Title	**Total active volume of customer´s deposits**
Abbreviation	VOLACT
What is being measured?	Sum of the total active volume of deposits of the employee´s customers
Formula (if applicable)	Sum of active volume of customer´s deposits per employee
Frequency	annual
Source	Controlling department

Table 17: performance measure record sheet (measure "Total active volume")

Except from the number of inhabitants within the catchment area, the data were provided by the bank´s controlling department. Since it was agreed that the present case study should be a test run, archival data from 2012 were used. Therefore, no conclusions about current performances of employees could be drawn. A total of 44 employees were working as account managers in 2012. Since four of them left the bank or changed their jobs within the bank during 2012, they were excluded from the sample. The 40 remaining employees all worked full-time. Thus, a full-time equivalent did not have to be calculated. The data were provided as linked-anonymous data using numbers from 1 to 40 instead of names. Although the data included personal data (e.g. salary) no sensitive data as defined in the Data Protection Act of 1998 were used. The data were transferred to an Excel spreadsheet.

In designing the DEA application, adequate DEA models had to be selected for each transformation stage. In a first step, this involved considerations about the rationale behind the performance evaluation. It was agreed that, in general, the purpose of the evaluation was not to reduce resource usage (inputs), but merely to stimulate improvement and growth of outputs. Additionally, it was assumed, that outputs could be changed or controlled by the employees more easily than inputs. Therefore, it seemed reasonable to employ an output-oriented model. Thus, the difference between a score of 1.00 and the score calculated represents the proportions by which the employees could improve their outputs by maintaining the same input. A second consideration was to decide whether to assume constant or variable returns to scale (see 2.3.2). The production of a banking service is, due to customer participation and a high degree of intangibility, very heterogeneous. Thus, the outputs of this process rely to a high degree on individual effort (e.g. to selling financial products to a customer) requiring the same effort each time the service is produced. In this respect, scale effects cannot be

anticipated. Thus, assuming constant returns to scale, the CCR model with an output orientation was employed for all stages[8].

High correlations among input and output variables may cause difficulties in DEA analysis. Thus, one purpose of the study was to find a "parsimonious" model. Adding highly correlated variables may result in the increase of performance scores, indicating that many DMUs are efficient, when they actually differ by only a small and random fluctuation. A parsimonious model typically shows low correlations among the variables and uses as many variables as needed, but as few as possible (Wagner et al., 2003). To ensure a parsimonious model, a correlation analysis for the set of proposed input and output variables was run for each stage. The results of the linear Pearson correlation (correlation coefficient r) are provided in the tables below.

	CUST	INHAB	VOLACT	SELL	APP
CUST	1.000	-0.280	-0.496	-0.097	0.379
INHAB		1.000	0.272	-0.030	-0.057
VOLACT			1.000	0.055	-0.061
SELLQ				1.000	0.305
APPQ					1.000

Table 18: correlation matrix TS1 "acquisition"

	SELL	APP	VOLACT	VOLL	BSH	RuV	Union
SELL	1.000	0.305	0.055	0.364	0.207	0.487	0.352
APP		1.000	-0.061	0.175	0.567	0.203	0.488
VOLACT			1.000	0.150	-0.267	0.195	-0.184
VOLL				1.000	0.499	**0.763**	0.604
BSH					1.000	0.410	0.533
RuV						1.000	0.598
Union							1.000

Table 19: TS2: correlation matrix TS2 "consultancy and sales"

	SAL	Bonus	TRAI	CMII	CommE
SAL	1.000	0.738	0.743	0.670	0.721
Bonus		1.000	0.686	0.397	0.542
TRAI			1.000	0.397	0.553
CMII				1.000	0.742
CommE					1.000

Table 20: correlation matrix TS3 "profitability"

[8] In TS4 „bonuses" an input-oriented CCR model was applied additionally as an alternative way to calculate the amount of bonus payments

	SAL	CMII	CommE	Bonus
SAL	1.000	0.670	0.721	0.738
CMII		1.000	0.742	0.409
CommE			1.000	0.542
Bonus				1.000

Table 21: correlation matrix TS4 "bonuses"

	CUST	INHAB	VOLACT	SAL	CMII	CommE
CUST	1.000	-0.280	-0.496	-0.290	-0.387	-0.474
INHAB		1.000	0.272	0.196	0.277	0.327
VOLACT			1.000	0.554	**0.778**	0.628
SAL				1.000	0.670	0.721
CMII					1.000	0.742
CommE						1.000

Table 22:correlation matrix TS5 "performance targets"

The higher the correlation coefficient r, the stronger is the relationship between the two variables. In literature, there are several categorizations concerning the strength of the relationship. A very common classification is the classification by Dancy and Reidy (2007):

Value of the Correlation Coefficient	Strength of Correlation
1	Perfect
0.7 - 0.9	Strong
0.4 - 0.6	Moderate
0.1 - 0.3	Weak
0	Zero

Table 23: classification of Pearson´s correlation coefficient r (Dancey and Reidy, 2007)

Other authors assess ranges from 0.8 to 0.9 as "very strong" or from 0.75 to 0.95 as "high" (Myers et al., 2010). Applying a cautious approach, all correlations that exhibited an r > 0.75 were marked as highly correlated. Examining the correlations with regard to this condition lead to several insights. Two variable combinations are highly correlated. The strongest relationship is observed between the variables "RuV" and "VOLL". Also, the variables "VOLACT" and "CMII" are highly correlated. All other variable combinations correlate on rather weak or moderate levels.

The correlation analysis was discussed with management. The correlation between "VOLACT" and "CMII" could be neglected since the variables belong to different categories ("VOLACT" serves as an input variable and "CMII" as an output variable). The

correlation of “VOLL” and “RuV” is between variables of the same category (outputs). However, management found both variables important for mapping the performance process precisely and for reflecting strategic directions of thrust. Thus, it was agreed not to remove them in a first run, but to investigate the issue in the context of a subsequent sensitivity analysis (which is discussed in section 6.3.2).

In a final stage of data collection, a data summary sheet was created for each stage. The sheet for the stage “acquisition” is illustrated below (the data summary sheets for TS2 to TS5 are provided in annex 1).

Data Summary

stage: acquisition

DEA model = CCR-O

No. of DMUs = 40
No. Input items = 3
Input(1) = CUST
Input(2) = INHAB
Input(3) = C VOLACT
No. of Output items = 2
Output(1) = SELL
Output(2) = APP

Returns to Scale = Constant (0 =< Sum of Lambda < Infinity)

Statistics on Input/Output Data

	CUST	INHAB	VOLACT (€)	SELL	APP
Min	111.90	2154.00	72395.50	1.00	65.00
Max	1648.50	39784.00	22319537.01	807.00	716.00
Mean	680.77	19081.13	2679019.41	160.00	342.98
SD	360.32	9936.90	5228208.20	116.89	149.98

Correlation

	CUST	INHAB	VOLACT	SELL	APP
CUST	1.000	-0.341	-0.496	-0.097	0.379
INHAB		1.000	0.273	-0.033	-0.120
CUSVOL			1.000	0.055	-0.061
SELLQ				1.000	0.305
APPQ					1.000

DMUs with inappropriate Data with respect to the chosen Model

No.	DMU
	None

Figure 33: data summary sheet TS1 “acquisition”

7.3 Performance Assessment

The following section presents the results for transformation stages 1 - 5 that were calculated applying the DEA CCR-O model. Additionally, a sensitivity analysis including a superefficiency analysis and weight restrictions was carried out to ensure the validity of results.

7.3.1 Results for transformation stages 1 - 5

Since multiple performance stages were assessed, performance scores were calculated separately for each transformation stage. To calculate the scores, a CCR-O model was used. No weight restrictions were incorporated. The calculation was computed with the Banxia Frontier Analyst Software (Version 4.2.0).

7.3.1.1 Results for TS1 "acquisition"

The table below holds the results for the transformation stage "acquisition".

Unit name	Score	Ref. count	number of peers	peers				Unit name	Score	Ref. count	number of peers	peers			
1	0.68	0	3	12	38	40		21	0.86	0	2	20	25		
2	1.00	1	0					22	0.52	0	4	12	34	37	38
3	0.66	0	4	36	37	38	39	23	0.48	0	3	12	38	40	
4	0.67	0	4	12	34	38	40	24	0.82	0	3	20	25	34	
5	0.67	0	2	20	34			25	1.00	11	0				
6	0.92	0	1	20				26	0.65	0	1	34			
7	0.37	0	2	34	36			27	0.41	0	4	12	34	37	38
8	0.53	0	1	40				28	0.93	0	1	34			
9	1.00	1	0					29	0.40	0	3	12	37	38	
10	0.91	0	2	20	25			30	0.90	0	3	25	38	39	
11	0.64	0	1	20				31	0.69	0	4	36	37	38	39
12	1.00	14	0					32	0.91	0	2	34	40		
13	0.99	0	2	12	20			33	0.31	0	3	12	25	37	
14	0.80	0	3	25	38	39		34	1.00	14	0				
15	0.46	0	4	25	36	37	39	35	0.38	0	3	12	34	40	
16	0.60	0	4	12	20	25	34	36	1.00	5	0				
17	0.78	0	3	12	20	25		37	1.00	8	0				
18	0.35	0	4	12	34	38	40	38	1.00	12	0				
19	0.78	0	4	12	20	25	34	39	1.00	6	0				
20	1.00	11	0					40	1.00	8	0				

Table 24: results for TS1 "acquisition"

In total, eleven employees achieved a score of 1.00. Those employees cannot improve their output, without increasing the input. In other words, if they had to acquire more customers they need to be assigned to another territory with more inhabitants or with customers that have higher average value in deposits. Those employees are also

referenced as peers to employees who obtained a score less than 1.00. One of the most valuable procedures in DEA is the comparison of an inefficient DMU with its counterpart – the peer or the reference set. The total reference count indicates how often an "efficient" employee serves as peer to others. The reference count does not equal the number of "inefficient" employees, but is slightly higher. This is due to the fact that an inefficient unit's reference set may contain several efficient units with a similar input/output orientation. Together they form a reference set for the inefficient DMU and provide examples of good operating practice for the inefficient DMU to emulate. This is why usually more than one peer is indicated per employee.

Considering the distribution of reference counts for TS1, it is striking that they are quite evenly distributed. Thus, nine out of eleven efficient DMUs repeatedly are assigned as references. This may be an indication, that the DMUs are distributed quite even on the efficient frontier and that there are not many outliers with very exceptional input-output structures. To verify this hypothesis, a subsequent sensitivity analysis should be instructive.

MIN	0.31
MAX	1.00
MEAN	0.75
SD	0.23
no eff. Units	11.00

Table 25: results summary for TS1 "acquisition"

The results summary for TS1 reveals that the range between the maximum and the minimum score is 0.69, which is quite large. The mean score however is within the upper third and the standard deviation of 0.23. is rather moderate. Thus, there may be some DMUs who are outliers downwards. Here, a more detailed look at the results on individual level should explain the result and give hints for improvement (see section 6.3.2).

7.3.1.2 Results for TS2 "consultancy and sales"

The results for the second stage "consultancy and sales" are presented in the tables below:

Unit name	Score	reference count	number of peers	peers			
1	0.35	0	5	7	19	26	29
2	0.64	0	4	7	19	29	35
3	0.53	0	4	7	19	26	29
4	0.32	0	3	7	26	29	
5	0.99	0	2	26	29		
6	0.62	0	3	21	26	29	
7	1.00	27	0	7			
8	0.58	0	3	7	26	29	
9	0.17	0	3	7	26	29	
10	0.78	0	3	21	26	29	
11	0.81	0	4	19	21	26	29
12	0.31	0	4	7	19	29	35
13	0.32	0	4	7	19	26	29
14	0.67	0	3	7	26	29	
15	0.63	0	5	7	19	26	29
16	0.88	0	5	7	19	26	29
17	0.68	0	4	7	19	26	29
18	0.72	0	4	7	19	29	35
19	1.00	20	0	19			
20	0.30	0	3	21	26	29	

Unit name	Score	reference count	number of peers	peers			
21	1.00	5	0	21			
22	0.30	0	3	7	26	29	
23	0.25	0	4	7	26	29	35
24	0.94	0	2	26	29		
25	0.69	0	3	19	26	29	
26	1.00	30	0	26			
27	0.42	0	3	7	26	29	
28	0.63	0	2	19	26		
29	1.00	33	0	29			
30	0.59	0	3	7	19	29	
31	0.48	0	3	7	26	29	
32	0.21	0	4	7	19	26	29
33	0.54	0	4	7	19	26	29
34	0.24	0	3	7	19	26	
35	1.00	10	0	35			
36	0.33	0	4	7	19	26	29
37	0.77	0	3	7	19	29	
38	0.48	0	4	7	19	26	29
39	0.44	0	3	7	26	29	
40	0.22	0	3	7	26	29	

Table 26: results for TS2 "consultancy and sales"

MIN	0.17
MAX	1.00
MEAN	0.59
SD	0.27
no eff. Units	6.00

Table 27: results summary for TS2 "consultancy and sales"

In contrast to "acquisition", the number of employees achieving the maximum score of 1.00 in "consultancy and sales" has decreased to six. Interestingly, none of the employees who managed to be efficient in stage "acquisition" is also efficient in consultancy and sales. Considering the six efficient employees in this stage, all of them serve as peers to others. However, employees no. 7, 26 and 29 are assigned as peers four times more often than the other efficient DMUs. This may be an indication that the set is dominated by those DMUs. The subsequent sensitivity analysis will shed more light on this. The range from maximum to minimum scores is 0.83 and with this even larger than in stage "acquisition".

7.3.1.3 Results for TS3 "profitability"

The tables below hold the results for the stage "profitability":

Unit name	Score	reference count	number of peers	peers		
1	0.34	0	3	5	10	16
2	0.49	0	2	5	10	
3	0.31	0	2	5	17	
4	0.28	0	2	5	17	
5	1.00	22	0			
6	1.00	23				
7	0.41	0	3	5	10	16
8	0.37	0	3	5	10	16
9	0.20	0	1	17		
10	1.00	19	0			
11	0.82	0	0			
12	0.74	0	3	10	16	25
13	0.91	0	2	16	17	
14	0.47	0	2	5	10	
15	0.70	0	2	5	17	
16	1.00	11	0			
17	1.00	0	0			
18	0.37	0	3	5	10	25
19	0.78	0	1	10		
20	0.57	0	2	5	10	

Unit name	Score	reference count	number of peers	peers		
21	0.83	0	3	5	10	16
22	0.18	0	1	25		
23	0.31	0	2	5	25	
24	0.93	0	2	5	10	
25	1.00	2	0			
26	0.85	0	2	5	10	
27	0.31	0	1	17		
28	0.47	0	1	10		
29	0.37	0	2	5	25	
30	0.36	0	2	5	10	
31	0.37	0	1	17		
32	0.48	0	2	16	17	
33	0.27	0	2	10	16	
34	0.73	0	1	10		
35	0.61	0	2	10	16	
36	0.56	0	2	5	17	
37	0.73	0	1	10		
38	0.47	0	1	17		
39	0.68	0	2	5	17	
40	0.34	0	2	5	17	

Table 28: results for TS3 "profitability"

MIN	0.18
MAX	1.00
MEAN	0.59
SD	0.26
no eff. Units	5.00

Table 29: results summary for TS3 "profitability"

Considering the stage "profitability", the number of efficient units decreases further down to five. Thus, there are very few employees who deploy input factors efficiently. Again, most of the efficient employees of this stage are not the ones who managed to be efficient in the previous two stages. This, on the other hand, indicates that employees with a superior performance in acquiring or consulting customers are not necessarily doing this efficiently. This issue will be addressed in the results review on organizational level. The range of scores and the mean score are similar to the previous stage "consultancy and sales".

7.3.1.4 Results for TS4 "bonuses"

The tables below illustrate the results for the stage "bonuses".

Unit name	Score	reference count	number of peers	peers		
1	0.31	0	2	10	16	
2	0.49	0	2	5	10	
3	0.31	0	2	5	6	
4	0.28	0	2	5	6	
5	1.00	22	0			
6	1.00	18	0			
7	0.39	0	3	5	6	10
8	0.35	0	3	5	6	10
9	0.15	0	3	10	16	25
10	1.00	24	0			
11	0.82	0	2	5	6	
12	0.74	0	3	10	16	25
13	0.61	0	2	10	16	
14	0.47	0	2	5	10	
15	0.70	0	2	5	6	
16	1.00	10	0			
17	0.74	0	3	10	16	25
18	0.37	0	3	5	6	10
19	0.78	0	1	10		
20	0.57	0	2	5	10	

Unit name	Score	reference count	number of peers	peers		
21	0.78	0	3	5	6	10
22	0.18	0	2	6	16	
23	0.31	0	2	5	6	
24	0.93	0	2	5	10	
25	1.00	6	0			
26	0.85	0	2	5	10	
27	0.28	0	3	6	10	25
28	0.47	0	1	10		
29	0.37	0	2	5	6	
30	0.36	0	2	5	10	
31	0.37	0	2	5	6	
32	0.33	0	3	10	16	25
33	0.27	0	2	10	16	
34	0.73	0	1	10		
35	0.61	0	2	10	16	
36	0.56	0	2	5	6	
37	0.73	0	1	10		
38	0.47	0	2	5	6	
39	0.68	0	2	5	6	
40	0.34	0	2	5	6	

Table 30: results for TS4 "bonuses"

	no restrictions
MIN	0.15
MAX	1.00
MEAN	0.57
SD	0.26
no eff. Units	5.00

Table 31: results summary for TS4" bonuses"

Like before, the number of five efficient DMU is rather small. The large range of 0.85 and the rather high standard deviation indicate a broad spread in the distribution of scores within the sample. For the stage "bonuses", the scores are not directly used to assess performance, but rather indirectly to calculate a performance oriented share of bonus. The rationale behind this procedure is to assess, in a first step, what revenues the employees achieved given their salaries and last year´s bonuses as input variables. In a second step, management needs to determine an overall amount of bonus, they which

to assign this year. Subsequently, in a third step the individual scores are applied to calculate the employee´s individual share of the overall bonus.

There are two alternative ways, to calculate the individual bonus share. Alternative 1 assumes the entire overall bonus volume (OB) will be distributed. To calculate the individual bonus for a specific employee(j) (= DMU(j) (j= 1, ….n)), the value of a score of 1.00 (SB) needs to be calculated in a first step by dividing the overall bonus volume through the sum of all bonus scores:

$$SB = \frac{OB}{\sum_{n}^{1} \theta} \qquad 1.18$$

Subsequently, the Bonus of DMU(j) (B(j)) is determined by multiplying the value of (SB) with the individual score of DMU(j) (θ(j)):

$$B(j) = SB * \theta_j \qquad 1.19$$

Alternative 2 assumes that the individual bonus share is a share of the average bonus per employee. This way, the overall bonus volume will not be exhausted, except all employees achieve a score of 1.00. In a first step, the average bonus volume per employee (AB) has to be determined by dividing the overall bonus volume (OB) through the number of employees (n) within the sample:

$$AB = \frac{OB}{n} \qquad 1.20$$

To calculate B(j), an input-oriented CCR model is employed. Thus, the average of this year´s bonus *AB* is added as an input measure for each DMU (since the value of *AB* is the same for each DMU, the individual score θ_j that was calculated for alternative 1 does not change).

TS4: bonuses (alternative 2)			
	Title	**Abbrevation**	**What is being measured**
Inputs	**Salary**	SAL	Salary
	Bonus	Bonus	Sum of bonus payments received
	Bonus expected	Bonus exp.	Average bonus share to be expected this year
Outputs	**Contribution margin II**	CMII	Contribution margin II resulting from interest earnings
	Commission earnings	CommE	Sum of commissions earned

Table 32: performance measures TS 4 (alternative 2)

To calculate the bonuses for those employees who received scores below 1.00, DEA provides rich information. Since DEA assesses the relative efficiency of a DMU compared to all other DMUs within the set, it not only calculates an overall performance score for each DMU, but also provides a projection of input-output levels, that would render a DMU efficient (Thanassoulis and Dyson, 1992). As a result of the input-oriented view, DEA projects the necessary “input reductions” for each input measure in for each inefficient DMU to become efficient. Since the average bonus volume was included as an input measure, B(j) results from the projected value for the input measure “bonus expected”.

Both alternatives were applied assuming an overall bonus volume of 77,000 € which management intended to assign for the account managers at the end of the year. The table below holds the results for both alternatives using the performance scores from TS4:

DMU no.	Score	Alternative 1 (CCR-O)	Alternative 2 (CCR-I)
1	0.31	1,051.87 €	414.69 €
2	0.49	1,666.02 €	652.91 €
3	0.31	1,051.87 €	508.41 €
4	0.28	939.89 €	372.02 €
5	1.00	3,393.12 €	1,925.00 €
6	1.00	3,393.12 €	1,925.00 €
...	...	...	...

Overall Bonus Volume (OB):	77,000.00 €	
Average Bonus (AB)		1,925.00 €
Sum of Scores:	22.693	
Value/Score 1.00 (SB)	3,393.12 €	
Sum overall Bonus Alternative 2		36,114.56 €

Table 33: excerpt of results of bonus calculation (alternative 1 and 2)

As expected, by using alternative 1 the average volume of the individual bonus (B(j)) is higher since the overall bonus is distributed entirely. An additionally appeal of this approach is that employees, whose performance is above average, also receive a bonus payment above the average bonus level. On the contrary, alternative 2 provides individual bonuses that have smaller volumes (e.g. with a performance score of 1.00 an employee receives 3,393.12 € using alternative 1 and 1,925.00 € using alternative 2). The advantage in using alternative 2 is that not the entire bonus volume is distributed. Thus the remaining volume can be used for other purposes, e.g. adding it to next year´s overall bonus as an incentive or handing it out to employees, who excelled in other areas (like social engagement). Additionally, since not the entire volume is distributed, the overall bonus volume can be increased over the years, which may also be an incentive to employees.

7.3.1.5 Results for TS5 "performance targets"

Finally, the results for determining performance targets are presented in the tables below.

Unit name	Score	reference count	number of peers	peers			
1	0.46	0	3	25	28	34	
2	1.00	4	0				
3	0.55	0	2	34	36		
4	0.42	0	2	34	36		
5	1.00	9	0				
6	0.73	0	2	5	10		
7	0.65	0	4	2	25	36	39
8	0.68	0	1	34			
9	0.58	0	3	2	25	28	
10	1.00	10	0				
11	0.61	0	4	5	10	25	28
12	0.79	0	2	16	34		
13	0.74	0	3	10	16	28	
14	0.73	0	3	5	25	39	
15	0.90	0	2	5	25		
16	1.00	6	0				
17	0.73	0	3	10	16	25	
18	0.56	0	2	34	36		
19	0.99	0	4	10	16	25	28
20	0.73	0	4	5	10	25	28

Unit name	Score	reference count	number of peers	peers			
21	0.70	0	2	5	10		
22	0.28	0	3	25	34	36	
23	0.46	0	1	34			
24	0.96	0	4	5	10	26	28
25	1.00	16	0	25			
26	1.00	2	0	26			
27	0.34	0	3	25	34	36	
28	1.00	11	0				
29	0.40	0	2	5	25		
30	0.66	0	3	2	25	28	
31	0.60	0	2	34	36		
32	0.54	0	2	28	34		
33	0.32	0	3	10	16	25	
34	1.00	15	0				
35	0.88	0	2	28	34		
36	1.00	9	0	36			
37	1.00	1	0				
38	0.86	0	3	25	34	36	
39	1.00	3	0				
40	0.50	0	1	34			

Table 34: results for TS5 performance targets

MIN	0.28
MAX	1.00
MEAN	0.73
SD	0.23
no eff. Units	11.00

Table 35: results summary for TS5 "performance targets"

Determining performance targets is an HR-practice that has long been applied in the bank. As already noted, both employees and supervisors were not happy with the procedure so far. According to statements given by the workers´ council and the bank´s management, this was mainly because performance targets had to be determined by "gut-feeling" or by simply looking up the employee´s last year results and add a certain percentage. Neither socio-personal circumstances like training or salary nor external influences like inhabitants in the catchment area were taken into account. Thus, the whole procedure was perceived as unfair by the employees and as an unpleasant task by most supervisors.

After a discussion with management and the workers´ council, it was decided to stick to the same targets (outputs) as before, but to take personal and environmental circumstances (inputs) into account. The results show, that eleven employees achieved a score of 1.00. Thus, the targets of those employees cannot be increased without also increasing some of the inputs. This could be a rise in salary (promotion) or the

assignment to another (larger) territory. To ensure that the efficient DMUs really performed exceptionally rather than being “just an outlier”, a subsequent sensitivity analysis of the results is crucial (Wagner et al., 2003).

The mean score is 0.73, which is quite high in comparison to the stages “profitability” and “bonuses”. The standard deviation is rather moderate. Thus, the overall performance concerning the achievement of targets is quite good and homogeneous. As already described for calculating bonus payments (alternative 2), DEA´s projections of input-output levels (“targets”) that would render a DMU efficient are used to construct achievable targets for those employees who received scores below 1.00. The table below presents an extract of those projections:

DMU no.	Score	actual		projection	
		CMII	CommE	CMII	CommE
1	0.46	55768.45	20024.08	122188.38	43872.66
2	1.00	108125.49	34319.65	108125.49	34319.65
3	0.55	67911.92	8139.46	123946.48	18193.46
4	0.42	38616.88	5784.05	91380.01	29480.8
5	1.00	395736.06	40797.61	395736.06	40797.61

Table 36: projection of performance targets (extract)

Although those projections give valuable information to construct achievable performance targets, they should be treated cautious. DEA literature often points out, efficiency may not always be achievable for each DMU, as other variables, that were not considered, may be important (N. Adler et al., 2002; Wang and Chin, 2010). Thus, the projections should rather be considered as maximum targets. In this context Paradi and Schaffnit (2004) suggested that, in the end, it should be up to management´s judgement based on their experience to assess whether the targets obtained by DEA are achievable or not.

7.3.2 Sensitivity analysis

A common issue in DEA is that it does not discriminate between the efficient units (see chapter 2.4). Furthermore, it bases its efficiency estimates on a comparison of input-output levels of an individual DMU with those of a very small subset of efficient peers. In this respect, it can be highly sensitive to data swings at the individual DMU level (Thanassoulis, 1993). A frequently noted advantage of DEA is that it assumes complete

substitutability of inputs and outputs and therefore assigns individual weights to derive the highest performance score possible for each DMU. An issue that may arise from such an analysis is that some DMUs may be evaluated highly efficient, although they excel in some of the determined performance criteria and at the same time assign zero weights to others (R. R. Thomas et al., 1998).

Since the results of the DEA assessment shall be used for evaluating employee performance on a continuous basis, the acceptance, and thus the validity of the data, is of crucial importance. Therefore, a sensitivity analysis of the results was performed before processing them any further.

7.3.2.1 Parameters for sensitivity analysis

For sensitivity analysis three issues that are often discussed in DEA application (see chapters 2.4 and 2.5) were taken into account:

Unrestricted assignment of weights

A first important parameter for sensitivity analysis concerns issues that arise from the unrestricted assignment of weights. Besides the situation that employees may excel in some criteria only, it may occur, that weights assigned by DEA are not reflective of management´s strategic map (R. R. Thomas et al., 1998). On the contrary, the so called "hands-off" policy of DEA is frequently mentioned as a major advantage by both researchers and practitioners. After a discussion with the bank´s management it was agreed that the "hands-off" policy should generally be retained. However, it was decided to rerun a weight-restricted DEA for each stage to see whether there are "specialists" (employees whose high rating is based on only few criteria). It was agreed to use a rather moderate restriction with $u,v \geq 0.1$. That is, each criterion had to be considered at least 10 percentage points in the calculation of the performance score. To assess whether the weights imposed had a significant impact on the efficiency rating of a certain DMU, Thanassoulis (1995) suggested to highlight cases were efficiency drops by 10 percentage points. As bank´s management and workers´ council felt this was too restrictive, it was decided to highlight cases were efficiency drops by 30 percentage points.

Detection of data errors

To detect data errors, it was agreed to remove one of the highly correlated variables ($r > 0.75$) and see whether the removal had a significant impact on the overall results. According to common statistical practice, a significance level of $p \leq 0.05$ was determined. Thus, the impact would be assessed as significant, if the mean score dropped or increased by more than 5 percentage points. If there were more than two highly correlated variables the procedure needs to be repeated accordingly.

Identification of outliers

To discriminate between efficient units and identify outliers, several methods have been proposed in literature. For the case-study it was decided to apply the concept of superefficiency for classifying efficient units for two reasons (see chapter 2.4). First, the Anderson-Peterson model is more convenient to calculate as there is no cross tabulation to create. Second, the results are easier to interpret for practitioners. Since the superperformance score of a particular DMU measures how much the efficient frontier is shifted towards the origin by the removal of that DMU, superefficiency can be interpret as a measure of the DMU´s influence (Wagner et al., 2003). To assess whether a DMU is of high influence, Wagner, Shimshak et al. propose a limit of 2.00, which was adapted for the case study. In a second step, the actual influence is investigated by removing DMUs with a score larger than 2.00. This is done by removing only one DMU for each DEA run (e.g. if DMUs A and B have a score greater than 2.00, one DEA is run with only DMU A removed, and a second with only DMU b removed). Again, a significance level of $p \leq 0.05$ was agreed to assess the impact of the DMU´s removal on the mean score. Thus, if the mean score drops or increases by more than five percentage points, the DMUs influence is significant. Although employees with a high superperformance score shall not be penalized for showing superior performance, it was agreed to remove them from the data set, if their influence was significant. The decision was based on the belief that otherwise it would be rather hard for all other employees to achieve efficiency.

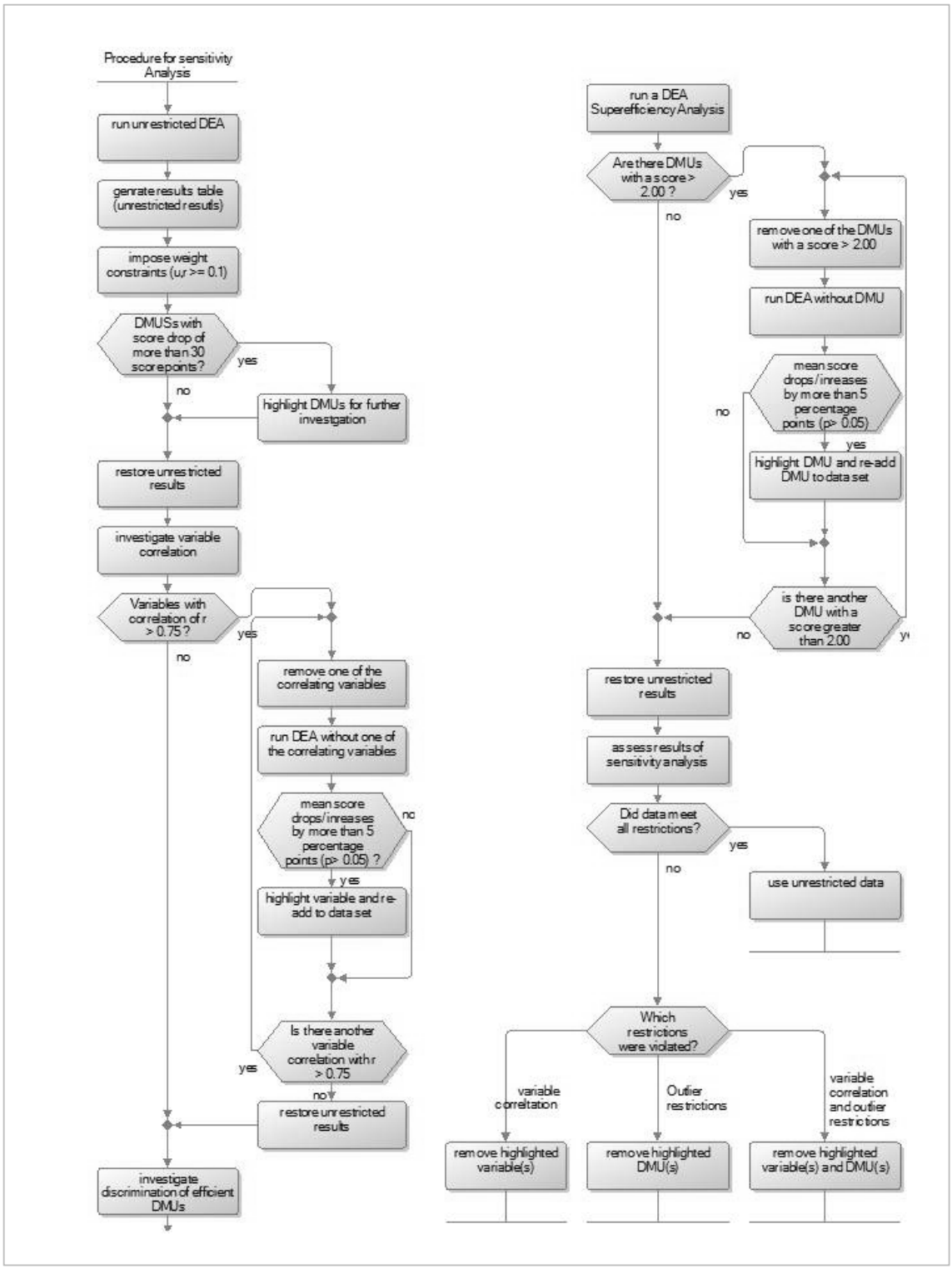


Figure 34: Procedure for sensitivity analysis

Figure 35 displays the whole process of sensitivity analysis in detail. To evaluate the influence of each of the mentioned factors separately and independently, it is of crucial importance to restore the data set after each DEA run, as illustrated above.

The procedure was applied to stages 1 to 5.

7.3.2.2 Application of sensitivity analysis

The application and the results of the sensitivity analyses are illustrated in the following. The detailed results for each stage are provided in annex 2.

The table below holds the summary of results from the sensitivity analysis for TS1.

	no restrictions	weights imposed (all inputs and outputs min 10%)	superefficiency	DMU removed (DMU 34)
MIN	0.31	0.01	0.31	0.31
MAX	1.00	1.00	3.11	1.00
MEAN	0.75	0.69	0.86	0.78
SD	0.23	0.25	0.48	0.23
no eff. Units	11.00	9.00	11.00	12.00

Table 37: sensitivity analysis - results summary TS 1 "acquisition"

In a first analysis, weights were imposed, ensuring that the weight for each variable in the calculation had to be by at least ten per cent. In doing so, the mean score drops from 0.75 to 0.69. At the same time the number of efficient units decreases from eleven to nine units. This shows that, in general, DMUs benefit from DEA´s unconstraint distribution of weights. Whilst most employees show only slight drops in scores or scores remain unchanged, employee no. 29 suffers a dramatic decline in score when weights are imposed. The employee´s score drops by 0.39 points down to 0.01 points. This is an important hint, that the employee may employ only one or few of the variables under consideration. This suggestion is confirmed by a closer look at the detailed results for employee no. 29.

	score	IO Cont CUST	IO Cont INHAB	IO Cont VOLACT	IO Cont SELL	IO Cont APP
unrestricted	0.40	26.5	29.7	43.8	0	100
weights	0.01	10	10	80	10	90

Table 38: Input-Output contribution employee no. 29 (TS1)

The input-output contribution reveals that employee no. 29 only employs the output variable "appointments" to achieve a performance score of 0.79. Assigning zero weights to selling, this variable remains unconsidered. By imposing weight constraints, the results of selling are considered with at least 10 percentage points. Now, employee no. 29 achieves only a score of 0.01, exhibiting superior performance in making appointments, but subsequently fails in selling financial products. Although the bank´s

management decided to stick to the “hands-off” policy of DEA, analysing data sensitivity by imposing weights delivers valuable information for training needs. Thus, employee no. 29 may achieve a much better result for TS1 in the next period by improving selling skills only slightly. The analysis also shows, that nine out of eleven employees, who achieved a score of 1.00 without weight constraints remain efficient after imposing weights (DMU 40 only suffers a drop of 0.01 score points. DMU 12 a drop of 0.13 score points). Hence, most efficient units employ balanced weights and their results are quite robust.

To investigate the discrimination of efficient units, a superefficiency analysis was run in a next step. One DMU, namely employee no. 34, achieved a score greater than 2.00 (3.11). Considering that employee no. 34 is the most cited reference in the unrestricted DEA run (14 citations), it becomes even more evident, that this employee is of high influence. However, the DMU was removed from the set and the calculation was re-run. The results confirm the DMU´s high influence since the mean score increases from 0.75 to 0.78. Since DEA investigates relative efficiency and compares all DMUs within the data set to one another, the remaining DMUs benefit from the removal of employee no. 34 since their relative performance scores slightly increase for the most part. Also the number of efficient DMUs increases up to 12. However, with a rise in score of 4 percentage points, the influence of DMU no. 34 is not significant within a p-level of 0.05. Hence, the DMU was not removed from the set.

In summary, sensitivity analysis for TS1 “acquisition” has led to the following conclusions. First, the unconstraint assignment of weights did not result in an unbalanced use of variables, with the exception of employee no. 29. In this case, the performance analysis on individual level needs to address this issue. Second, there were no highly correlated variables in the set. Therefore, all variables remain within the data set. Finally, it was recognized that within the set of eleven efficient DMUs, the performance of employee no. 34 is of high influence. Since the removal of this particular DMU led not to a significant change in mean score, the DMU remains within the set. Consequently, the unrestricted data set can be applied for further investigation.

Considering the stage “consultancy and sales” the results from the sensitivity analysis draw a different picture:

	no restrictions	weights imposed (all inputs and outputs min 10%)	variable removed (RuV)	super-efficiency	DMU removed (no. 26)	DMU removed (no. 29)	DMU removed (no. 7)	DMUs (no.7, 26 and 29) and variable (RuV) removed
MIN	0.17	0.01	0.13	0.17	0.19	0.17	0.27	0.34
MAX	1.00	1.00	1.00	10.00	1.00	1.00	1.00	1.00
MEAN	0.59	0.25	0.56	1.08	0.63	0.70	0.63	0.77
SD	0.27	0.22	0.26	2.00	0.28	0.26	0.25	0.23
no eff. Units	6.00	2.00	5.00	0.00	10.00	8.00	5.00	11.00

Table 39: sensitivity analysis - results summary TS 2 "consultancy and sales"

Thus, the sensitivity analysis for TS2 brings to light some interesting peculiarities. By imposing weights, the means score drops considerably from 0.61 to 0.25. Thus, it seems that a majority of employees employ only some of the variables and assign zero weights to others. This is confirmed by a closer look at the results on employee level. In summary, 23 out of 40 DMU suffer a loss of more than 30 score points if weights are imposed. However, when taking a look at the input-output distribution of these DMU, there is no evidence that there are particular variables that are being "avoided". Rather, it appears that performance is very heterogeneous considering the output variables. Due to the rather large number of outputs, each employee has some leeway to achieve the best score possible.

In terms of variable correlation, the outputs "RuV" and "VOLL" have an r-value of 0.763. It was decided to remove "RuV" from the data set, since the variable is just one out of three that represents a particular financial product. "VOLL" on the other hand represents the total volume of liabilities handled by the employee and therefore was evaluated as being more holistic by the bank´s management. With "RuV" removed, the mean score drops to 0.57, which is more than five percentage points from the unrestricted mean.

The subsequent superefficiency analysis detects three highly influential DMUs. With scores of 2.54 (DMU 7) 8.94 (DMU 26) and 10.00 (DMU 29), they achieve extremely high values. Interestingly, DMUs 26 and 29 are the only ones who remain efficient after weights are imposed. Additionally, all three DMUs are frequently assigned as peers to all other DMUs in the set. This leads to the assumption that they achieve their outstanding results by outstanding performance rather than by a rare input-output structure. Removing DMU 7 leads to a rise in mean score up to 0.63, removing DMU 26 also lets the mean score rise up to 0.63, the removal of DMU 29 even leads to a rise up to 0.70. Thus, all DMUs are of significant influence and dominate the set. Although results suggest that the employees no. 7, 26 and 29 are no outliers in the sense of operating with a rare input-output-structure (since all three are among the most cited references), they dominate the remaining DMUs significantly. Since performance evaluation should also address motivational aspects, it may not be wise to appoint peers

or calculate individual performance targets mainly based on an outstanding performance. Thus, it was decided to remove the three DMUs from the data set for TS2. Nonetheless, these employees receive a score of 1.00 for this stage. FoTS2, sensitivity analysis results in the removal of the variable “RuV” and in the removal of DMUs 7, 26 and 29, which lead to the following results.

For TS 3 “profitability”, sensitivity analysis led to the removal of DMU 6 from the set. For the stages “bonuses” and “performance targets” no restrictions were imposed.

7.3.2.3 Summary and Results correlation

After sensitivity analysis, the data sets of TS2 “consultancy and sales” and TS3 “profitability” had to be modified in order to meet the defined restrictions. The tables below hold an overview of the results that were subsequently used for further analysis and for deriving HR-strategies on organizational and individual level.

stage	TS1: acquisition	TS2: consultancy and sales	TS3: profitability	TS4: bonuses	TS5: performance targets
Restriction	none	DMUs (no.7, 26 and 29) and variable (RuV) removed	DMU (no. 6) removed	none	none
Unit name	Score	Score	Score	Score	Score
1	0.68	0.61	0.34	0.31	0.46
2	1.00	0.75	0.49	0.49	1.00
3	0.66	0.76	0.40	0.31	0.55
4	0.67	0.45	0.46	0.28	0.42
5	0.67	1.00	1.00	1.00	1.00
6	0.92	0.62	1.00	1.00	0.73
7	0.37	1.00	0.41	0.39	0.65
8	0.53	0.85	0.38	0.35	0.68
9	1.00	0.41	0.22	0.15	0.58
10	0.91	1.00	1.00	1.00	1.00
11	0.64	1.00	1.00	0.82	0.61
12	1.00	0.50	0.74	0.74	0.79
13	0.99	0.67	0.91	0.61	0.74
14	0.80	1.00	0.47	0.47	0.73
15	0.46	0.92	0.88	0.70	0.90
16	0.60	1.00	1.00	1.00	1.00
17	0.78	0.83	1.00	0.74	0.73
18	0.35	1.00	0.38	0.37	0.56
19	0.78	1.00	0.78	0.78	0.99
20	1.00	0.45	0.57	0.57	0.73
21	0.86	1.00	0.84	0.78	0.70
22	0.52	0.39	0.23	0.18	0.28
23	0.48	0.41	0.34	0.31	0.46
24	0.82	1.00	0.93	0.93	0.96
25	1.00	0.98	1.00	1.00	1.00
26	0.65	1.00	0.85	0.85	1.00
27	0.41	0.76	0.49	0.28	0.34
28	0.93	1.00	0.47	0.47	1.00
29	0.40	1.00	0.41	0.37	0.40
30	0.90	0.85	0.36	0.36	0.66
31	0.69	0.86	0.65	0.37	0.60
32	0.91	0.36	0.48	0.33	0.54
33	0.31	0.90	0.27	0.27	0.32
34	1.00	0.41	0.73	0.73	1.00
35	0.38	1.00	0.61	0.61	0.88
36	1.00	0.69	0.63	0.56	1.00
37	1.00	0.84	0.73	0.73	1.00
38	1.00	0.97	0.86	0.47	0.86
39	1.00	0.86	0.76	0.68	1.00
40	1.00	0.34	0.57	0.34	0.50

Table 40: Overview results TS1 - TS5 after sensitivity analysis

	TS1: acquisition	TS2: consultancy and sales	TS3: profitability	TS4: bonuses	TS5: performance targets
MIN	0.31	0.34	0.22	0.15	0.28
MAX	1.00	1.00	1.00	1.00	1.00
MEAN	0.75	0.79	0.64	0.57	0.73
SD	0.23	0.23	0.25	0.26	0.23
no eff. Units	11.00	14.00	7.00	5.00	11.00

Table 41: summary results TS1 - TS5 after sensitivity analysis

Comparing the results before and after sensitivity analysis, it becomes evident that most employees benefit from the modifications. Thus, the mean scores did rise in all stages were data were modified according to the restrictions. So did the number of efficient units. Particularly in TS2 “consultancy and sales”, employees benefited from the modifications with an increase in mean score of twenty score points and an increase in efficient units from six units up to fourteen. Additionally, the range from minimum to maximum score was reduced considerably in stages TS2 and TS3. However, since DEA is a comparative approach, this does not imply that performance in those stages got better. Rather, dominating DMUs were removed from the set (mainly for motivational reasons) which made performance more homogeneous.

In table 51 the results for each stage are marked. The results marked “dark grey” label the efficient units. Results marked “light grey” label results that are above the mean score. Results below the mean score are not marked. By studying the marks more closely, it becomes apparent that individual performance varies considering the distinct stages. For example: employee no. 26 achieved a score of 0.65 in acquisition, which is below average. However, when evaluated in terms of consulting and sales, she shows excellent performance and also performs above average considering profitability.

Apparently, there is no employee, who manages to excel in all stages. On the other hand, only four employees show poor performance throughout all stages. Since employees tend to be more inclined to act on advice in areas of poor performance when there are also areas of good performance (see chapter 4.3), this may be of great advantage for counselling purposes. The initial attempt of determining distinct performance stages was to distinguish between the various requirements that are included in the job profile. Those requirements are not necessarily complementary. Thus, to investigate the relationships between the CCR scores of the five stages, a linear Pearson correlation coefficient was calculated. The table below holds the results.

	TS1: acquisition	TS2: consultancy and sales	TS3: profitability	TS4: bonuses	TS5: performance targets
TS1: acquisition	1.000	-0.297	0.334	0.291	0.494
TS2: consultancy and sales		1.000	0.380	0.436	0.412
TS3: profitability			1.000	0.915	0.656
TS4: bonuses				1.000	0.750
TS5: performance targets					1.000

Table 42: correlation analysis of results TS1 - TS5

First, it should be noted that, except for two cases, there are only very weak to moderate correlations between the results (< 0.75). This indicates that employee´s performance in one stage generally is rather independent from the other stages. Again, this provides evidence for the necessity of a multi-dimensional approach.

With an r-value of -0.297 there is a weak negative correlation between the stages “acquisition” and “consultancy and sales”. Thus, employees who show good performance in acquiring customers tend to be rather unsuccessful in consultancy and sales, which is an issue that should be investigated further. The stage “profitability” shows moderate correlations to the stages “acquisition” and “consultancy and sales” but a very strong correlation to the stage “bonuses”. Thus, employees who manage to achieve good performance without wasting inputs tend to benefit from bonus payments more often than employees who excel in other stages. Since improving profitability was defined as number one strategic target by the bank´s management, followed by the increase of sales volume (consultancy and sales) and the enhancement and enlargement of customer contact (acquisition), the distribution of bonuses adequately reflects the hierarchy of strategic targets. This is an indicator for DEA´s suitability to implement and cascade strategic targets into HR-practices.

After sensitivity analysis, the results are analysed and discussed considering two perspectives: the individual and the organizational level.

7.4 Performance Review

Reviewing the performance evaluation results, there are two general perspectives:

a) Deriving recommendations to improve competences and performance (developmental purpose) and deciding about promotion, performance targets or bonuses (administrative purpose) on individual level

b) Deriving norm strategies for Human Resources Management on organizational level.

The first perspective, looking at the individual results, may be more useful to line or branch managers. The other perspective, drawing conclusions affecting the bank´s HR-strategy, is more geared towards senior management. Both perspectives are examined as follows.

7.4.1 Reviewing results on employee-level

To illustrate the scope and the level of detail of the results obtained, an exemplary review is provided for employee 6. A major issue of practical applications in DEA is that results sometimes fail to impress management due to the fact that they are presented in a manner that is too complex and hard to understand for practitioners (Paradi and Schaffnit, 2004). To present the results in a more convenient manner for managers to use them for communicating feedback, discussing strengths and weaknesses and to decide about performance targets, promotions or bonuses, a “performance sheet” was created for each employee. An excerpt is presented below:

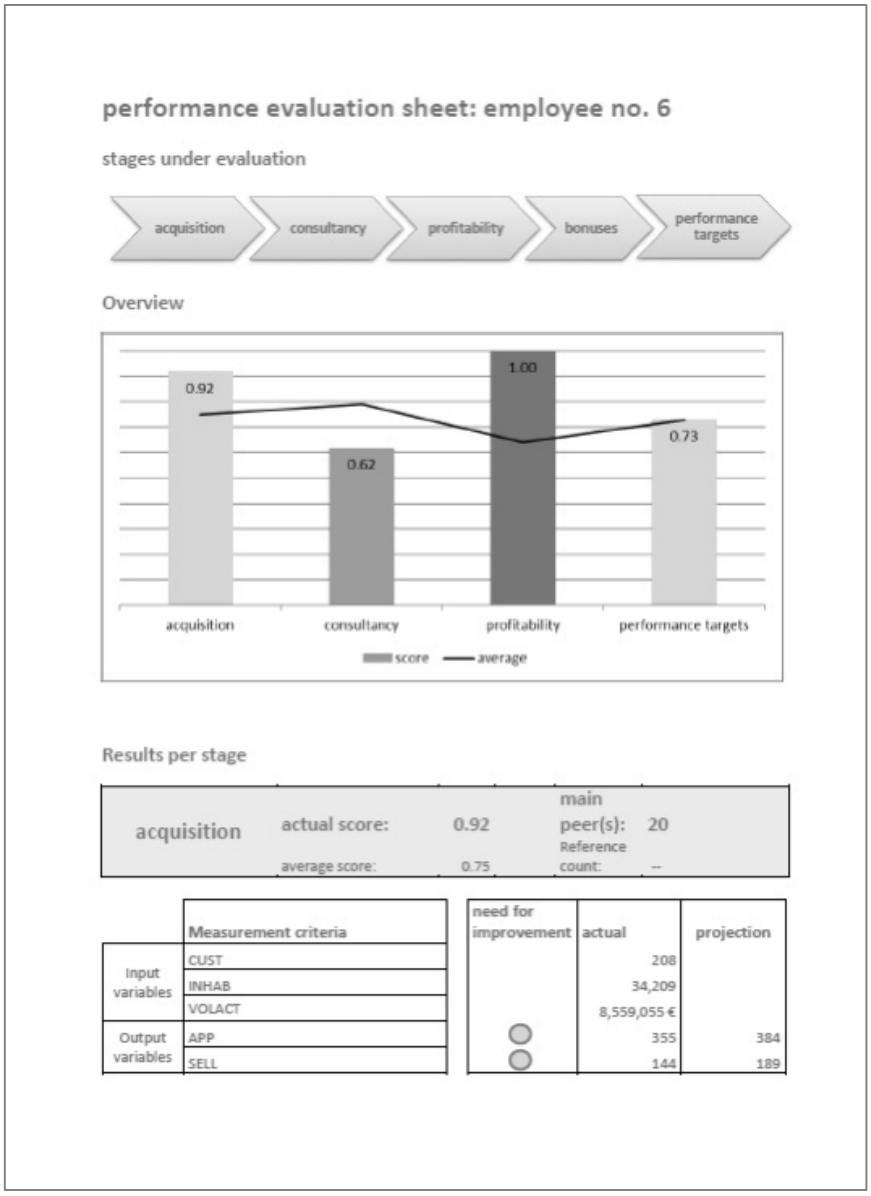


Figure 35: Performance Evaluation Sheet for employee no. 6 (excerpt)

The results sheet is divided into three sections. Section 1 illustrates the performance stages assessed. Section 2 holds a graphical overview about the results the employee obtained for each stage and the average score (the stage “bonuses” is not included, since the score obtained is only used to calculate the bonus). Thus, section 2 gives a rough sketch of the employee´s performance without holding too much detail. It may be used for management summaries or as a starting point for appraisal interviews. Section 3 presents the results per stage in more detail. It holds information about the assigned peers (or, in case of a score of 1.00, about the frequency the employee is cited as reference). Furthermore, all input and output variables are listed. Considering the output, the actual values are compared to target values (“projections”). A traffic light system indicates whether there is no need for improvement (actual equals target value), a

moderate need for improvement (difference between actual and target value is less than 50 percentage points) or a strong need for improvement (difference between actual and target value is more than 50 percentage points). The ranges for the traffic light system were determined in cooperation with management.

Although DEA provides more details than the information that is presented in the "performance sheet", the bank´s management reported back that the information is sufficient for the defined purposes (see chapter 8). If more information is needed (e.g. other peers than the main peer, modification in score if weights are imposed etc.), the managers are additionally provided with the complete results table. The full "performance evaluation sheet" for employees 6 is provided in annex 3.

Employee no. 6 exhibits a quite satisfactory performance in acquiring customers. With a score of 0.92 the results are slightly above average. Looking at the output of this stage more closely, employee no. 6 is far better in making appointments with customers than in selling products. Whilst the potential improvement for making appointments is very moderate, the potential improvement for selling quota is about 30 percentage points. In TS2, the score drops below average. A closer examination reveals that employee no. 6 takes major advantage of the output variable "active volume" with an output contribution of 95 percentage points. Thus, by imposing weighting constraints, the score drops dramatically down to 0.01. Since "active volume" represents the total volume of customer´s deposits per employee, this variable is rather a consequence from selling financial products than a financial product itself. At this point it should be examined, whether the employee acquired most of her customers herself (and with this achieved the good results in "active volume" herself) or if she took them over from another colleague. In the latter case (and under consideration of the results in TS1), there is a need for training concerning the employee´s selling skills. Besides a common sales training, it could also be a good idea to team her up with one of her peers (e.g. employee no. 21 who has the highest lambda-value).

Regarding the stage "profitability", the employee shows superior performance. Although she has shortcomings in selling financial products, she generates high revenues in relation to her labour costs. Thus, the kind of products the employee actually sells, generate very high contribution margins. In relation, her labour costs are quite low, as employee no. 6 did not have any training expenses or receive a bonus the year before. As a consequence, expenses for a sales training seem more than reasonable. Additionally, the employee receives the highest possible bonus as a result of the superior performance score. This could also serve as a motivation to improve her selling skills.

Improved selling skills may also help to achieve the performance targets, which aim at a very moderate increase of both contribution margin and commissions earnings.

7.4.2 Reviewing results on organizational level

On organizational level, the DEA results should be a reliable basis to derive conclusions concerning the strategic alignment of the bank´s HR-practices. In order to derive strategic decisions, a two-dimensional strategy portfolios may provide some assistance. The figure below shows a portfolio combing the stages “acquisition” and "consulting and sales”.

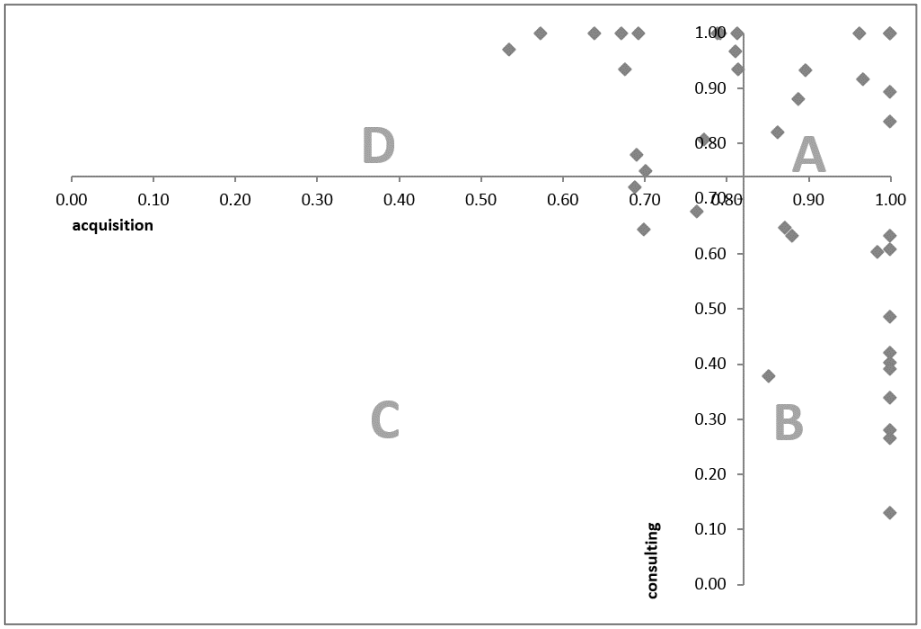


Figure 36: acquisition-consultancy portfolio

The portfolio is divided into four fields with the dividing lines drawn with respect to the average performance of the two stages under consideration. A comprehensive analysis enables management to derive norm strategies for each of the four strategic fields (Koch-Rogge et al., 2014). By implementing strategies to move employees closer to field A, a movement of axes should also be observed (if the overall performance improves, the average score points will increase).

The portfolio graphically displays that there are many employees who show superior performance in one of the two stages. Whilst the employees are allocated quite even in fields A, B and D, only a minority of employees is located in field C. With respect to HRM activities, management should discuss norm strategies how to reinforce the bank´s ability to attract new customers by maintaining the strengths of consulting and sales at the same time. A closer examination of the four fields can be a good basis for this.

- Field A (field population: 12): The employees in this field show above-average performance in both stages. Although this field is among the most populated fields, there is a lack of “superior” performance (those units would be found in the top right corner). This was already indicated by the correlation analysis of results. The matrix shows that it seems to be difficult for employees to excel in both stages. If a closer analysis of the individual results confirms the impossibility to excel in both stages, the appraisal interview should be used to inform employees which stage to emphasize according to the organization´s strategy.
- Field B (field population: 10): Those employees are successful in acquiring new customers or arrange appointments with existing customers, but show below-average performance in generating revenues from those appointments. An obvious strategy in this case is to enforce sales trainings. A different strategy could be to team those employees up with employees that are good in sales and counselling (field D). That also seems reasonable in a practical sense, since both fields are nearly equally populated. A strategy that requests organizational restructuring – but is nonetheless worth discussing - is a segregation of functions into “canvassers” and “consulters”. This way, each employee could focus on her strengths.
- Field C (field population: 6): The employees in field C exhibit a rather poor performance in both acquisition and consulting. This group certainly needs the most attention and supervision from management. To identify the reasons for poor performance, a closer look at the results on individual level should be helpful. Since the field is populated with six employees only, the additional effort is kept in limits. In general, there are different norm strategies conceivable. Again, the enhancement of training may be a suitable strategy, especially since four employees perform just “a little below average” in acquisition or consultancy and sales respectively. Thus, slight improvements may be sufficient to move them to another field. Another strategy is to assign them to new territories (e.g. with another customer structure) or another branch, so that the change of environmental factors may be a basis to improve their performance. If performance continues to deteriorate, another job profile may be considered.
- Field D (field population: 12): Contrary to the employees in field B, employees in field D fail to acquire customers, but they are above-average performers concerning. Equal to the strategies named for field B, training or teamwork may improve their performance.

Similar portfolios can be created for investigating the relationships between the other stages. Over time, employees may move from one field to another due to altered performance. Thus, it is important to monitor how each employee changes in performance relative to all other employees, whose performance may also change. As a consequence of these changes in performance, modified HR strategies may be requested (Westermann, Johnson 1999). An even more comprehensive insight can be gained by adding a third stage to the three-dimensional analysis. In order to avoid a three-dimensional portfolio, the results are presented in a three-digit approach. Columns 2-4 indicate a high (above average) or poor (below average) performance per stage. Subsequently, column 5 indicates the allocation of employees across the respective field.

field number	TS1: acquisition	TS2: consultancy and sales	TS3: profitability	field population
1	high	high	high	9
2	high	high	low	3
3	high	low	high	4
4	high	low	low	6
5	low	high	high	6
6	low	high	low	6
7	low	low	high	2
8	low	low	low	4

Table 43: three-dimensional view of strategic fields

The holistic analysis of three stages confirms the impression that there is quite a large spread in total performance since all fields are populated. Apparently, the majority of employees are situated in field one. This is pleasing for the bank´s HR management since nearly ten percent of the employees under consideration show above average performance in three stages. Fields 4,5, and 6 are evenly populated. Whilst employees in field 4 excel in acquisition but lack consulting and sales skills by an additional waste of inputs, employees in field 5 exhibit a quite opposite performance. This strongly supports the idea of an organizational restructuring with the segregation of functions into “canvassers” and “consulters”. Alternatively, to improve performance in both fields without changing the bank´s organizational structure, a working environment should be created in which both parties can learn from one another or work in teams.

Furthermore, the analysis reveals that only a minority of two employees show below-average performance in all three dimensions (field 8). This confirms the assumption that it is very likely that the employees perform well in at least one stage, where they may

even be referenced as a peer to others. Considering motivation and willingness to improvement, this is an important issue. Additionally, the importance of taking into account multiple performance stages in order to assess employee performance becomes evident. Not only does this procedure grasp the whole performance process, but it also enables management to derive accurate strategies that align with the organization´s strategic goals (Koch-Rogge et al., 2014).

7.5 Conclusions from the case study

Regarding the application of DEA on employee level several conclusions on considering its technical suitability can be drawn from the case study.

Considering the performance assessment by DEA, its application in the case study showed, that the unrestricted DEA run in general gave reasonable results, identifying good and weak performers, indicating areas and levels of possible improvement and indicating peers as best-practice benchmarks. At the same time, the sensitivity analysis has proven to be important to evaluate results from the unrestricted DEA regarding their relevance, robustness and plausibility. Thus, to identify an unbalanced use of inputs, to prevent performance targets that are too demanding or to identify employees with outstanding performance, a subsequent sensitivity analysis proved to be crucial to adjust the results to the organizational needs. Additionally, the results proved that the vast majority of employees performed above average in at least one stage, which may have a high impact on acceptance of results and willingness to improve. To get an encompassing impression of the employee´s performance it would have been desirable to include the customer´s view at some points. Especially the stage “consultancy and sales” would have profited from including a variable concerning customer satisfaction.

In terms of reviewing individual performance based on DEA´s results, the “performance evaluation sheet” was developed. This sheet allows for a customized review for each employee. By including many details on individual employees´ performance it may serve both developmental and administrative purposes. At the same time, it aims at reducing complexity by presenting the information in a user-friendly dashboard-style. Further, the aggregation of individual results by performance portfolios may inform HR decisions on organizational level. By this, DEA may provide a linkage between the individual and organizational perspective.

8 Assessing the Applicability of DEA for measuring Employee Performance

For a Performance Evaluation System to unfold its full potential it needs to be accepted and understood by all stakeholders of the evaluation process. In addition, it needs to provide reasonable results that are suitable for being used as the basis for HRM activities. Therefore, the assessment of the proposed DEA-based approach by all relevant stakeholders is presented in this chapter. Since the stakeholders form a rather heterogeneous group and the assessment should provide both in-depth qualitative statements as well as quantitative empirical findings, the assessment is based on both qualitative and quantitative data.

Initially, the assessment by managers and workers´ council is presented regarding methodology, data collection and findings. Respectively, this is also reported for the subsequent assessment by employees. Finally, the assessment´s findings are summarized.

8.1 Assessment by management and workers´ council

The assessment among managers and workers´ council on whether DEA meets the organizational requirements to a performance evaluation on employee level was carried out by conducting several Focus Group interviews. The purpose of this assessment along with the choice and justification of the methodology were discussed in chapter 6.2.3.3. Hence, this chapter illustrates the planning of the focus group interviews including the design and sampling process. Further, the analysis of results using Thematic Analysis (TA) and the discussion of results are presented.

8.1.1 Planning and Conducting the Focus Group interviews

In planning Focus Group interviews several determinants need to be considered including the study´s design and sampling, the size and segmentation of the groups, the interview content and the moderator´s level of involvement (Kitzinger, 1995). In the following sections, the planning phase of the Focus Group interviews for this research is outlined.

8.1.1.1 Design and Sampling

In a first step, potential participants needed to be identified and acquired. Since the proposed method was applied in one particular bank only (that hereafter is referred to as "Bank A") the bank´s management and the members of the bank´s workers´ council were considered most important for participating in the Focus Group assessment. They had already agreed in advance to participate in the subsequent assessment. Further, to address some of the concerns considering the validity and reliability of single-case findings, the assessment was expanded to groups of managers and workers´ councils from other regional cooperative banks (see chapter 6.2.2.4). Therefore, the management and workers´ council from seven other banks (with similar characteristics concerning size, geographical location, number of branches and clients) were contacted and asked to participate in the assessment. Eventually, managers and members of workers´ council from three of the contacted banks (which hereafter are referred to as "Bank B", "Bank C" and "Bank D") agreed to take part in the study. In addition, to receive feedback from an overarching level, the associations of both German Savings banks and German Cooperative banks were asked to participate in the Survey. The association of German Savings banks agreed to take part in the study represented by HR managers of their regional sub-association: the association of savings banks of Lower Saxony.

Whilst some researchers claim that Focus Group studies can consist of six to over fifty groups, others state that there should be more than just one single group (Kitzinger, 1995). According to Kitzinger (1994), Focus Group interviews are conducted until theoretical satisfaction is reached. Kruger and Casey (2015) suggest to plan for three or four Focus Groups with a particular type of participant. If an adequate saturation has not been reached then, additional interviews should be conducted. Hence, the number of groups becomes larger the more heterogeneous the groups are. To provide rich data, the members of the Focus Groups need to fully engage in the discussion and therefore should feel comfortable with each other. For this reason, most researchers advocate the use of homogeneous groups (Rabiee, 2004). On the other hand, bringing together a diverse group of participants may promote the exchange of different perspectives within a group setting (Kitzinger, 1994).

Crucial characteristics for forming a group are: age-range, social class, gender, and ethnic group (Freitas et al., 1998). Besides the level of homogeneity, the number of participants in each group is a crucial issue. One important principle is that the group should be small enough for each participant to have the opportunity to share her perceptions. At the same time, it should be big enough to enable a diversity of perceptions (Krueger and Casey, 2015). A moderate size of four to ten people is a

commonly agreed scope (Freitas et al., 1998; Kitzinger, 1995). Another important issue considering the design of the Focus Group study is whether the analysis shall be carried out as a single category design or as multiple category design. The latter one is conducted with different types of participants and allows the researcher to make comparisons among different groups (Krueger and Casey, 2015).

For the assessment at hand, the distinction between managers and workers´ council is crucial since it seems reasonable that the distinct groups may have different perceptions considering the performance evaluation they shall assess. In addition, Kitzinger (1995) points out that different hierarchy levels within the group may prevent a vivid exchange and even may affect the data. Hence, the author believes that the hierarchical level and the participants´ role within the organization were the most important characteristics for forming the Focus Groups. Accordingly, the Focus Group interviews followed a multi category design. Thus, five groups of consisting of the banks´ management (including chief executives, board members, HR managers, line mangers) and two groups of workers´ council[9] (including chairmen and members) were formed.

Group ID	sample size	role within organization		female (male)
A-M	9	chief executives	2	2 (7)
		board members	2	
		HR managers	1	
		line managers	4	
B-M	10	chief executives	2	4 (6)
		board members	2	
		HR managers	3	
		line managers	3	
C-M	7	chief executives	1	2 (5)
		board members	0	
		HR managers	2	
		line managers	4	
D-M	6	chief executives	1	3 (3)
		board members	0	
		HR managers	2	
		line managers	3	
S-M	3	chief executives	1	1 (2)
		board members	0	
		HR managers	4	
		line managers	0	
A-W	5	chairmen	1	1 (4)
		members	4	
B-W	6	chairmen	1	2 (4)
		members	5	

Table 44: Sampling of the Focus Group Assessment

[9] Since the workers´ council of Banks C and D consisted of less than three members each and bringing them together was not practicable, no Focus Groups could be formed in those banks.

The size of the groups ranged from 3 to 10 participants. When adding the members of the Savings Banks Association to the managers´ groups, there were five groups to evaluate the proposed method from a managerial perspective. Only two groups represented the workers´ council´s point of view. Thus, the author acknowledges that this may be a limitation for this study.

8.1.1.2 Topics and Questions

Since the purpose of the Focus Group interviews was to assess the technical and organizational suitability of the proposes DEA method for evaluating employee performance, the study was mainly of descriptive nature. Therefore, a set of pre-defined questions and sub-questions was provided in a topic guide list (which is provided in annex 4). The questions were determined in coherence with the conceptual framework of this research and the associated research questions. Thus, the guide list included the following key questions:

- How does the proposed method meet general requirements (perceived fairness, understandable, non-biased)?
- How does it serve administrative purposes?
- How does it serve developmental purposes?
- How does it meet the legal requirements of the organisation?
- How does it compare to traditional performance evaluation methods?

To investigate different experiences across the distinct groups of stakeholders (and also between the distinct group members) questions concerning personal experience with performance evaluation and experience considering other methods were included as introductory and ending questions.

According to Krueger and Casey (2015), questions in a focus group interview should be easy to say, clear, short and open-ended. These suggestions were followed for all key questions. However, some sub-questions were formulated as closed questions, which enhanced the comparability of the answers across groups. To evoke a conversation, rather than just the answering of questions, the moderator was encouraged to ask for further explanations, descriptions or illustrations by using “how” and “why” questions.

Following a question-route, the first questions on the topic guide list were opening questions, which were intended to bring each participant to talk early in the discussion (Krueger and Casey, 2015). These were followed by some introductory questions to

connect the participants with the topic (e.g. personal involvement in the evaluation procedure). After the introductory part, the five key questions were discussed. Eventually, the ending questions reflected on further suggestions for the evaluation procedure and on concluding remarks from each participant.

8.1.1.3 Moderation

Depending on the moderator´s level of involvement, the order of a topic list could be followed strictly or rather serve as an orientation. In case of a low-level-involvement the moderator introduces the first topic followed by an unstructured discussion until the next topic is introduced. This way, the discussion may reveal new and unexpected issues and therefore supports exploratory research. Furthermore, the moderator is able to observe the participant´s behaviours and to evaluate their interests. On the other hand, this kind of discussion is hard to control. Thus, the groups are relatively disorganized and important topics may be missed. Additionally, the results are more difficult to analyse. If the moderator is highly involved in the discussion, she controls the order and dynamic of the topics discussed. In this context, the topic guide list is followed rather strictly. Hence, the high-involvement strategy may be preferred for descriptive research. Whilst this approach makes sure that all desired topics are covered, there is the risk that the moderator´s bias will produce biased data (Kitzinger, 1995; Rabiee, 2004).

Since the purpose of the Focus Group interviews for this study were mainly of descriptive nature, but also had some exploratory elements (e.g. to identify themes for the subsequent questionnaire study), the author decided to follow an approach of “moderate involvement”. Thus, in the opening and introductory part there was minimal intervention from the moderator, which allowed the participants to set priorities (Kitzinger, 1994). Later on, the moderator adopted a more interventionist style by ensuring that all key questions were discussed, although not necessarily in the order of the topic guide list. Further, the moderator evoked the debate by asking for participants for clarification and to elucidate their point of view. Since the Focus Group interviews were not recorded due to the wish of the participants (see also chapter 6.2.3.3 and 6.2.5), the author did not facilitate the Focus Group discussions herself but rather attended to take notes and to observe. Thus, an independent moderator, who is familiar with DEA and has experience in the banking sector, conducted the sessions.

8.1.1.4 Interview Procedure

Initially, one session was scheduled for each group. Since group A-M engaged in a vivid discussion, a further session was scheduled for this group. Eight sessions were conducted in summary. The sessions were conducted in the period from December 2013 to April 2014 and lasted between two and three hours. This is longer than the duration recommended in literature, which is one to two hours (Rabiee, 2004). The author acknowledges that this may have a negative impact on the participants´ ability to focus and pay attention to the discussion. Therefore, a break was usually made after one and a half hour. Further, the participants were informed about the anticipated duration and agreed beforehand. The sessions were conducted in meeting rooms on site. The meeting rooms of all sessions provided a relaxed and calm atmosphere. Another advantage over meeting in office rooms was, that the sessions were not interrupted by phone calls etc. In addition, most rooms had workshop equipment including a beamer and flip charts.

At the beginning of each session the moderator explained the aim of the focus group interview to the participants. They were encouraged to talk to each other rather than to the moderator. In order to introduce the topic, the proposed method was illustrated by a 25-minute presentation at the beginning. In the presentation, the practical and scientific background of the method was pointed out, the basics of the DEA methodology were outlined and the method´s practical application in Bank A was illustrated. Subsequently, an exemplary performance evaluation sheet was explained and analysed. After the presentation, the participants were given the opportunity to ask questions.

To facilitate the beginning of the discussion and initiate a vivid talk, the participants were asked about their previous experiences with performance evaluation. In the further course of the session, the moderator paid attention to guide the discussion along the topic guide list. However, the moderator did not force the discussion into a certain direction. Thus, some key questions were not discussed chronologically according to the topic guide list. Eventually, the sessions ended with a concluding summary in which the moderator asked the participants what they considered most important during the discussion. Based on those statements, he gave a short overview on the most relevant points of the discussion.

Each session was documented by the author by taking notes. Subsequently the notes were transferred into interview protocols, which were send to the participants afterwards for review. It should be noted that all sessions took place in German language since all participants and the author were native speakers. Therefore, the notes and initial

protocols that were provided to the participants also are in German. For further analysis, the coding and the final reports were carried out in English.

The documentation procedure is illustrated in more detail in the following section.

8.1.1.5 Documentation

In general, Focus Group sessions are audio recorded and transcribed afterwards. In doing so, the author can reproduce all spoken words and sounds, including hesitations etc., which provides rich information for further analysis (Braun et al., 2012). Hence, audio recordings provide the highest level of details (Krueger and Casey, 2015). As already outlined before (see chapter 6.2.3.3 and 6.2.5) the participating banks did not wish the sessions to be audio recorded. This was mainly due to the fact, that it is not very common in Germany to audio-record meetings, especially in the banking sector which relies on a confidential atmosphere. Moreover, chief executives feared potential consequences under employment law even if all participants agreed to recording.

According to Krueger and Casey (2015), not all studies require the highest level of detail for quality analysis. In particular if they are not mainly of exploratory nature (Rabiee, 2004). Thus, some (rather descriptive) studies can be highly effective with much less detail. Therefore, the author decided to document the sessions by taking notes and to transcribe them into interview protocols afterwards. The author is aware that by taking notes, not all details of the interviews could be captured and a line-by-line analysis was not possible. This issue has already been discussed when considering the limitations of the research methodology in chapter 6.2.6.

For documenting a Focus Group session by taking field notes, Krueger and Casey (2015) suggest two distinct approaches:

a) To capture notes and quotes

b) To capture the discussion word for word as far as possible.

Whilst the latter option holds the advantage of providing more details and rich descriptive information, it places high demands on the skills of the person taking notes and may lead to missing out essential parts of the conversation. Therefore, the author decided to capture key points and findings of the conversation by notes. In addition, insightful quotes were captured word by word if possible.

In preparing the sessions, the questions and sub-questions from the topic guide list were numbered, so that the author could refer comments and notes to specific questions. To identify the speakers, the name of each participant was coded using the initial letters of

their name and surname. Further, date, duration, location, number and role of the participants were included in the first page of the notes. The notes sheet itself was parted into three sections: question number, notes, quotes. In addition, the moderator documented the concluding summary at the end of each session on flip chart. These notes supplemented the notes taken during the sessions.

After each session, the author conducted a 15- to 30-minute debriefing with the moderator and thereafter transcribed all notes into an interview protocol. Notes were transcribed with only minor changes that had no influence on the content (including cutting out comments that had nothing to do with the research whatsoever). After transcription, the protocols were sent to all participants along with the request to verify the content and give further comments if they felt something they said were misinterpreted. All 46 participants answered and verified the protocols. 15 of them indicating minor modifications mainly concerning the wording of terms, which were eventually incorporated into the protocols. A sample page of an interview protocol is provided in annex 6.[10]

8.1.2 Data Analysis

To establish a trail of evidence and to enhance the extent of consistency, data collected from Focus Group sessions should be analysed in a systematic, sequential, continuous and verifiable way (Krueger and Casey, 2015). For this research, the Thematic Analysis (TA) by Braun and Clarke (2006) was applied to analyse the gathered data. The selection and justification of this method has already been discussed in chapter 6.2.3.3. In the following sections, the analysis process is illustrated.

8.1.2.1 Coding

Since qualitative research, and Focus Group discussions in particular, creates vast amounts of data, a structured data analysis consists of several steps (Rabiee, 2004). According to the structure of TA, a first crucial step for every researcher is getting familiar with the collected data (Braun et al., 2012). For this research, an important phase for familiarizing with the data was transcribing the notes into protocols and subsequently incorporating the feedback from the participants into the data. When the protocols were finalized, the researcher eventually re-read all of them.

[10] The original notes were in German language. The sample page provided in annex 6 was translated into English.

The TA method assumes that, in a next step, initial codes are developed to identify and provide labels for the data that are of potential interest for the research. The coding can be carried out as a top-down process (deductive) with the researcher bringing a series of topics or concepts to the data, or as a bottom-up process (inductive) deriving the codes from what is within the data (Braun et al., 2012). Krueger and Casey (2015) suggest that the purpose of the study should drive the analysis. Since the purpose of this analysis is of mainly descriptive, but also holds some exploratory elements, the coding was done by using a combination of both approaches. First, a basic code list was created based on the literature review and the research questions. For each code, several sub-codes were defined. After the first Focus Group sessions with groups A-M and A-W, the code list was tested and enhanced to allow the development of a comprehensive code list. The test and review of the initial list took place after the interview sessions with the Focus Groups from Bank A, since the author assumed that these sessions would provide the richest information.

Supplementing codes to the initial list was done in a rather inductive manner, creating codes from the protocols that were based on the author´s interpretation of the pattern. In consequence, some initial codes were merged, renamed or deleted from the list. At this point, it should be noted that the author is aware that the coding process is a subjective selection to a certain extent. The final code list was then applied to code all protocols from the Focus Group sessions. The full code list and an excerpt of a coded protocol are provided in annexes 5 and 7.

8.1.2.2 Identification of Definition of Themes

After the data is fully coded, the analysis should move on to identify themes. According to Braun and Clarke (2006), a theme represents a level of patterned response within the data set that is related to the research questions. They also state that themes usually do not “emerge” from the data as sometimes referenced by other researchers. Rather, the construction of themes is an active process. TA suggest three distinct steps to create themes: identification, review and naming of themes.

For the first step, the coded data was reviewed regarding the identification of patterns, similarities or overlaps between the codes. Thereafter, codes that shared a unifying feature were clustered to first drafts of themes. Since the codes consisted of several sub-codes, some codes were associated to multiple themes. The first draft held nine distinct themes. In a second step, the themes were reviewed checking for quality of the theme, for the boundaries of the theme and whether there is enough data to support the theme (Braun et al., 2012). During this process, some themes were merged or their boundaries

were extended. Further, the theme on strengths and weaknesses was discarded and relocated under several other themes. To support the creation process, the clustering of codes to themes was illustrated in a map.

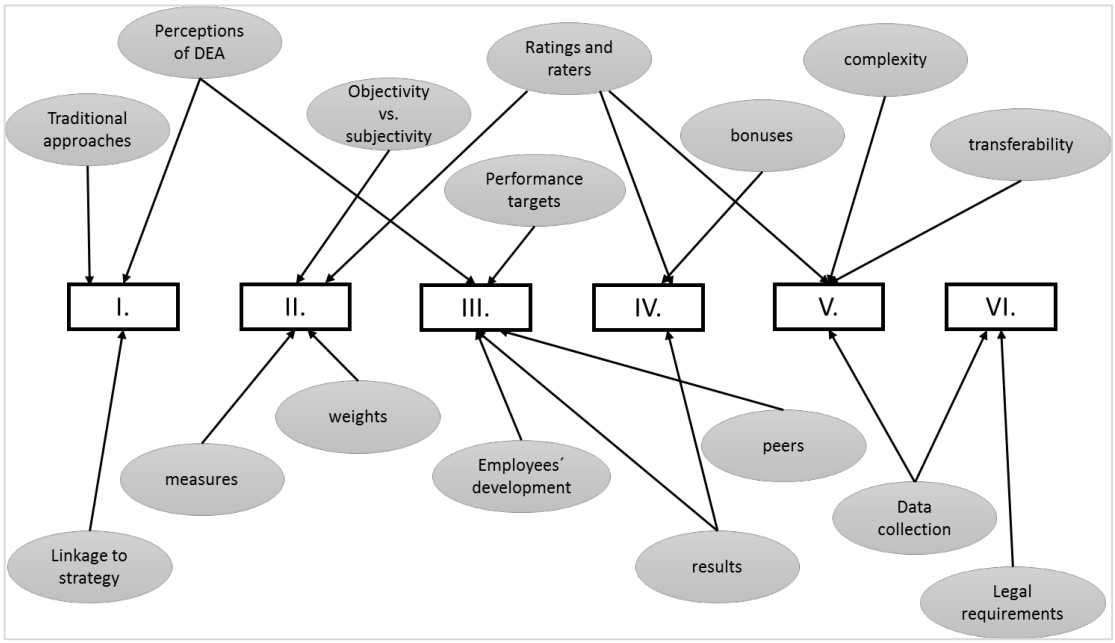


Figure 37: Themes map

In a last step, the identified themes were checked for not overlapping, each addressing a single and meaningful focus and contributing to answer the research question. Further, the coded protocols were revisited to make sure that all codes were covered by themes. Eventually the themes were labelled trying to sum up the essence of each theme in a few words. Thus, from the analysis the following six themes were generated:

I: Traditional methods and the DEA method

II: Objective measures

III: Managing employees´ development

IV: Transfer, promotion and bonuses

V: Challenges for implementation

VI: Compliance with legal requirements

The following chapter reports on the results of this analysis.

8.1.3 Results

The results of the data analyses presented in this section are discussed addressing each theme separately. In the analysis it was found that there was no apparent difference in views of managers´ groups and groups of the workers´ council. This is why the results for both are reported together. However, different opinions across groups or individual opinions, which did not necessarily reflect the group consensus, are highlighted accordingly. Quotes are highlighted in italics.

I: Traditional methods and the DEA method

The participants across all groups had at least some experience with performance evaluation. Besides, all banks employed traditional methods to measure employee performance. Those ranged from ranking systems (including several financial figures) to annual appraisal interviews (based on assessment sheets completed by line managers). There was widespread agreement throughout all groups that their current methods for assessing employee performance are not appropriate for a comprehensive evaluation and that they lack significance considering administrative or developmental issues. The group from the Savings Bank Association confirmed this impression:

> *"[...] Most regional banks lack knowledge and an appropriate system for evaluating their employees´ performance and are aware of this problem. Especially when it comes to combining employee performance and remuneration, methods are often not properly designed. [...]"*

Also the HR managers from banks B and D pointed out that their current evaluations methods are flawed:

> *"[...] Our current system lacks an input-oriented view. Thus, it does not account for environmental conditions or work experience. [...]" (Bank B)*

> *"[...] We conduct appraisal interviews once a year. Most of the managers perceive them as an additional task that does not add much value. [...]" (Bank D)*

The groups consisting of members from the workers´ council were even more critical considering the method their banks currently employed. They claimed that the current evaluation methods were often subject to complaints mostly because the employees felt treated unfair or were given unachievable goals. The proposed DEA-based method was received very positive by most participants across all groups. This was regardless of their level of familiarity with the method. In particular, the method received the highest approvals by the groups from bank A, who provided the case study data.

The majority of participants agreed that the proposed method was very suitable for translating business strategy into individual goals. Group A-W praised the breakdown of the whole performance process into several performance stages. A member of the group stated

> *"[…] it clearly shows that an employee´s efforts can serve more than one purpose. […]"*

A manager from group C-M found:

> *"[…]I think it is a good way to review performance from different angles. Besides, there are only very few employees who perform poorly in all stages. […]"*

Moreover, all groups agreed that the presented results were way more extensive and detailed than results provided by traditional methods they know and therefore most participants perceived the proposed method as superior to the methods they currently apply.

II: Objective measures

All participants emphasized the importance of choosing measures appropriately. They emphasized that a balanced use of measures is very important for a fair appraisal. Managers from bank A reported that, although they had most measures readily available, the process of identifying the "right" measures was rather hard but crucial. One participant of group A-M acknowledged the importance to take the time to define and review appropriate measures:

> *"[…] if we only had one workshop for defining measures, we would have ended up with more than thirty distinct measures. A second workshop to reduce the number and to see which measures were important and available was crucial. […]"*

She pointed out that if the measures were chosen poorly, the whole evaluation may be flawed.

A participant of Group C-M suggested that it should be considered beforehand which measures are readily available when determining measures. This suggestion was received rather ambivalent. Other members from this particular group pointed out, that all relevant measures that affect performance need to be included. A decision on measures should not only be based on their accessibility. However, there was broad agreement in all groups that, due to the method´s complexity, the PES should be conducted no more often than once a year.

There was broad agreement across all groups that the method is highly objective and free from bias. However, an issue that was discussed controversially across groups was the "forced" assignment of weights. Most groups (including both groups from the workers´ council) valued the fact that the proposed approach calculates weights by mathematical optimization techniques and thus is independent from forced weights. They also feared that this would contradict the perceived fairness of the method.

> *"[…] With our current evaluation method, results are influenced whenever weights are assigned to the measures. The major advantage of this method is its independency from "forced weights. […]"*

On the other hand, participants across all managers´ groups were in favour of the possibility of setting weight restrictions in order to stress the importance of some measures. Most of the participants evaluated the method´s objectivity as advantageous and as a major prerequisite for being accepted by employees. Some groups debated whether a "subjective assessment" should also be part of the performance evaluation. One participant from group A-M argued in favour of a subjective component:

> *"[…] Although that would make the method vulnerable to bias, a subjective component should be included to get the full picture. […]"*

This issue was discussed controversial across all managers´ groups. The groups consisting of members of the workers´ council stressed the importance of the method being independent from managers´ bias. Associated with this issue, several participants suggested that "soft-skills" should also be considered in order to arrive at a comprehensive assessment.

III: Managing employees´ development

Most participants found the results very helpful for counselling employees considering training needs and career goals. They emphasized the benefit of not only rating employees´ performance, but also being able to point out areas employees need to improve and to identify circumstances under which they cannot improve further.

> *"[…] It is quite interesting to see those results. I think, for supervisors it is important to understand why someone achieved better results than others. […]"*

As a concluding statement, group B-W indicated that the proposed approach accounts for both, managing poor performance and recognizing good performance. Many participants showed great interest in the change of performance considering the distinct stages. They found it noticeable that in the case study there was only a small group of

employees who were good at both acquisition and sales. Moreover, the method´s ability to take into account that employee´s performance is based on different input-output combinations, and therefore not all are comparable to one another was perceived very positive across all groups. A manager from group B-W stated:

> *"[...] It is fascinating to see, that there is only a small group of employees that are good at both acquisition and sales. This is a strong argument for creating distinctive job profiles. [...]"*

Also, it was mainly agreed that the performance evaluation sheet is a very helpful tool to counsel employees considering their strengths and weaknesses. In addition, all groups suggested that the employees should receive an overview of their results in advance, so they can prepare for the appraisal interview. Considering performance targets, there was general approval for the way performance targets were calculated. However, some participants in different groups expressed concern that some of those projections were too demanding. This is especially the case for targets suggesting an improvement rate of more than 50 per cent. A participant of group A-M remarked:

> *"[...] Since the method is based on a mathematical approach, the results provided are based on linear optimization. Thus, the results need to be adjusted by "common sense". [...]"*

Also, the possibility of assigning peers was debated controversially across all groups. In general, most participants appreciated the possibility of assigning peers. A participant of group A-M argued:

> *"[...] This way, employees understand that the targets are not somewhat "made up" but are oriented towards the best performances within the sample. [...]"*

A participant of group B-M added:

> *"[...] For managers on the other hand, peers are very interesting. This information facilitates decisions about who to bring together in a team and by this an improved" learning environment" can be created. [...]"*

However, there were intense discussions on this issue in most groups, since there were also strong arguments against naming peers. Some participants strongly advocated that peers should not be revealed to the employees in order to prevent negative working atmosphere and a "negotiation of results". Others, especially the workers´ council groups, expressed concern about violating confidential data. They argued that peers need to agree in advance to be named.

IV: Transfer, promotion and bonuses

Several managers pointed out that the results support transfer decisions, since they indicate when employees are not able to improve further under the given circumstances. One participant emphasizes that "salary" has been used as an input factor in several stages:

> *"[...] This seems very wise since by including salary as a measure, the method also helps to review of promotion decisions. [...] "*

The groups consisting of workers´ council emphasized the importance of choosing the measures carefully if administrative decisions shall be derived from the results. Furthermore, participants of several groups suggested to monitor the effect of transfer decisions.

Across all groups there was approval for the way the method enabled the calculation of bonuses. Most participants agreed this to be a traceable and fair method to calculate bonus payments. Several groups suggested that bonuses should only be paid limited to a certain performance or a certain score. Members of the workers´ council emphasized that the bonuses for managers needed to be calculated in the same manner, elsewise the method would be perceived as unfair by the employees. In general, this view was shared by the managers´ groups. However, they pointed out that for evaluating managers´ performance, measures need to be modified. In this context, another participant of group A-M suggested to include a measure like "leadership quota", that reflects the executive´s amount of time spent on leadership.

V: Challenges for implementation

Across all groups, the proposes method was evaluated as superior compared to traditional methods the banks currently employ. Moreover, most participants stated that they would support the method´s implementation in their organization for evaluating employee performance. However, despite the method´s advantages over traditional methods, the participants also identified challenges for an implementation, in particular concerning the method´s transferability to other organizational units, the effort for data collection and its complexity.

Concerning the method´s transferability, some participants expressed concern whether the method would work for other groups of employees, in particular for employees working in the back office. As a manager of group A-M pointed out:

> *"[…] It is hard to develop criteria for employee´s working in the back office. Although we already defined some measures that would work for back office employees, they were finally dismissed for they were either not available or not concise. […]"*

Another participant of this group added that there are measures that are suitable to measure back office performance, but the data collection would cause additional effort. Other groups discussed this issue in a similar manner.

One participant pointed out that the method required a certain number of DMUs to provide reasonable results, which makes it less attractive for evaluating smaller groups. Further, it was debated in some groups that some measures that are relevant for performance are not quantifiable, for example the "ability to work in a team". This is particularly true for so called "soft skills". Thus, participants of several groups suggested combining the method with a rather traditional appraisal. Concerning the issue of data collection, managers of Bank A pointed out, that although most of the data were readily available, data collection was a very time consuming process. At the same time, they stressed the importance of taking the time to select the measures properly. Participants of both management and workers´ council groups expressed concern over the effort that has to be put into data collection. Others argued that most of the measures that were used in the case study are readily available in most banks.

It was mainly agreed that the proposed approach is quite complex and therefore not simple to understand. Some participants admitted that they got the general idea, but had not dealt with the mathematics behind it. A manager of group A-M stated:

> *"[…] I needed some time to get familiar with the method and with interpreting the results. However, I guess when you got behind the general principle it is not necessary to fully understand the mathematical background. […]"*

Concerning the general rationale of the method, all participants agreed that it is necessary to understand the major principles. In this context, several groups noted that the performance evaluation sheet is very helpful and concise and that the results are understandable once the method has been explained. Members of the workers´ council emphasized that the results should be explained very carefully to the employees. A participant from group B-W pointed:

> *"[…] It is important that managers fully understand the results and how they are calculated. Otherwise, there is a danger of misinterpreting or even misusing the results. […] "*

Thus, it was acknowledged across all groups that there will be additional costs for training managers to get familiar with the approach and to interpret the results. The vast majority of participants ranked the proposed approach superior to the methods they currently use. Nethertheless, many of them expressed concern over the method´s comprehensibility. Thus, some participants feared that the results could overstrain some employees and managers. A participant of group B-W emphasized:

> *"[…] It is crucial that the methodology and the results are made comprehensible to the employees. Otherwise, the results may be rejected. […]"*

A member of the workers´ council from group A-W stated:

> *"[…] The complexity of the results may overstrain some employees. Therefore, employees should be provided with the results before the appraisal interview. Furthermore, they should have the opportunity to consult a member of the workers´ council before the appraisal interview. […]"*

It was therefore agreed that this issue should be considered when assessing the proposed method among employees.

VI: Compliance with legal requirements

None of the groups expressed concern that the proposed method violates existing laws. However, they all emphasized that the guidelines of the "Remuneration Ordinance for Institutions", which came to force in 2013, apply. The group of the Savings Bank Association indicated that the remuneration ordinance requests to align remuneration with the organization´s strategy. They evaluated the proposed approach is an appropriate tool to execute those requirements.

Another law that needs to be considered is the German General Equal Treatments Act. In particular, the workers´ council groups pointed out that, according to this law, employees have to agree to be named as peers. Considering confidential issues, several groups expressed concern about spreading the results. Thus, some participants of group C-M claimed that not all managers should have access to all results. They suggested that access should be strictly limited to those who need to work with the results. At the same time, there was broad agreement that all employees should have full access to all their results. In this context, a member of group B-W pointed out that, if the DEA runs are carried out by an external company, data anonymity it must be ensured.

Across all groups there was broad agreement that the workers´ council should be involved in the evaluation process. However, there was some disagreement about the

degree of involvement. Most management groups advocated a rather “passive” involvement, meaning hearing the workers´ council on PES matters and involving them in defining and reviewing the measures. The workers´ council groups and the group of the Savings Bank Association on the other hand, recommended an “active” involvement of the workers´ council during the whole process.

8.2 Assessment by employees

Whilst the managerial perspective is commonly considered in the evaluation of performance measurement methods, there is a lack of empirical research considering employee´s perceptions and opinions (C. E. Pettijohn et al., 2001). The inclusion of employee´s perception of an HR practice is also a claim that stems from HRM research. Thus, many authors suggest that the perception of HR practices and employee´s response to it, may be a core element to opening the “black box” and by this explain the linkage between HR practices and organizational success (Wright and Nishii, 2007). Furthermore, the consideration of employees´ perception may add another dimension to the assessment since the appropriateness of the proposed approach may be best evaluated from the viewpoint of the individuals most affected by it – the employees themselves. Therefore, an empirical study based on a questionnaire survey was carried out among 122 employees working in the service sector. The following section reports on the methodology and provides a descriptive and explanatory analysis of the results.

8.2.1 Planning and conducting the questionnaire survey

The study´s underpinning methodology has been outlined in chapter 6.2.3.4. Thus, this chapter illustrates the questionnaire development, sample selection and the procedure of data collection.

8.2.1.1 Questionnaire Development

The primary focus of the study was to examine employees´ perception of the proposed approach for evaluating employee performance. The questionnaire consisted of 50 questions. Although the questions were based on the literature review and the Focus Group discussions, they were all developed and formulated by the author (Saunders et al., 2012). Following recommendations from Bryman (2015) and Saunders et al. (2012), questions were mostly posed as closed questions keeping open questions to a minimum of one question (remarks). Since the purpose of the study was of descriptive and explanatory nature, most questions were posed in rating format using a five-point Likert

scale (Bryman, 2015). The order of response categories was kept the same throughout the questionnaire to avoid confusing the respondents (Saunders et al., 2012). Saunders et al. (2012) also point out, that the numbers in the rating questions should reflect the respondent´s feelings. Since a 1 to 5 scheme represents the German grading system (1 = excellent performance), the scale was labelled from 1 (= strongly agree/ very suitable) to 5 (= strongly disagree/ very unsuitable) so that a low score indicated approval. Accordingly, the response categories were listed from positive to negative format. This was done to visually aid respondents´ interpretation. However, the author acknowledges that in doing so, there may be implications of bias. Since a 5-point Likert scale was not appropriate for comparing distinct methods, a four-stage scale (with (1 = better, 2 = worse, 3 = about the same, 4 = I do not know the method) was applied for section C.

The questionnaire was divided into six thematic sections. The introductory section did not focus on the proposed approach itself. Rather, the participants were asked to rate the importance of characteristics that were identified to be of importance for a performance evaluation by literature review (see chapter 4.2). This section was included for two reasons. First, in order to interpret the employees´ assessment, it is crucial to know which characteristics of a PES they perceive as most important. Second, based on this data, a subsequent analysis can be performed to investigate whether the employees´ assessment of the general requirements relates to their perception of the proposed approach.

The next section (“Section A: general requirements”) investigated whether the proposed approach met those criteria considering three dimensions: measuring employee performance, determining performance targets and calculating bonuses. In the following section (“Section B: administrative and developmental purpose) the participants were asked to assess the proposed approach considering the two distinct purposes of a PES.

The next section (“Section C: comparison to other methods) focused on how the proposed approach compares to other methods. The other methods named were derived from the results of the Focus Group Interviews. Considering the purpose of evaluation of employee performance”, the method was compared to simple ranking formats, written appraisals and the 360-degree feedback. For the purpose of determining performance targets, the method was compared to adding a certain percentage to the employee´s last year´s targets and to determining the same targets for everyone (both methodologies were applied by banks that took part in the Focus Group Interviews). Finally, the calculation of bonus payments was compared to determining bonuses by management decision or to distributing a fixed share of the corporate profit. For all three questions, participants were given the possibility to name and rate other methods they know.

In the next section (“Section: Overall assessment”) participants were asked to assess the approach´s general suitability for evaluating employee performance, determining performance targets and calculating bonuses. Furthermore, they were asked whether they would endorse the application of the proposed approach in their organization. The final section consisted of the demographic and biographic data of the respondents. The participants were provided with a questionnaire in German language. The English version can be found in annex 8.

8.2.1.2 Sample Selection and Sample Characteristics

Since the case study was based on data of 40 sales employees of Bank A, choosing those employees as sample would have been an obvious choice. However, there were several reasons that lead to the decision to select another sample. First, one purpose of the study was to enable a comprehensive empirical analysis, which required a larger sample size. Second, another study purpose was to reflect a general view of employees´ perceptions, which required identifying a sample that experienced various types of performance appraisals and that included various types of job profiles. Thus, using just the 40 sales employees of Bank A as a sample, the study´s results would have provided a “micro view” generalizable to only one organization or at maximum one sector. Third, there was a probability that the employees´ perception of the method would be influenced by their individual results (this is to say that someone who received good results would be more likely to rate the method more positive than someone with rather poor results).

Therefore, a stratified random sampling method was employed in which employees, who were engaged in job activities in the service sector, were identified and approached as potential candidates (Bryman, 2015). Respondents who met the criteria, were identified in a continuing Bachelor program and in seminars for Quality Management both offered by Hochschule Harz in Wernigerode. The socio-demographic background of the participants was quite diverse. Participants from the continuing Bachelor program all were engaged in studies of business administration parallel to employment. Mainly, they worked in the financial industry or in service units of the automotive industry in middle management or as administrators. Their highest level of educational attainment mostly was “Abitur” (general qualification for university entrance in Germany) and partly “Fach-Abitur” (vocational baccalaureate diploma). Participants from the Quality Management seminars, on the other hand, were not engaged in studies at Hochschule Harz. Moreover, they attended the seminars provided by Hochschule Harz as additional job training. Since the seminars focused on service quality in particular, most of the participants worked in the tourism industry, but also in the retail and public sector. The

participant´s educational background was very diverse, ranging from finished vocational training to an academic degree.

The participants were contacted by the author and asked to participate in the study. Overall, 146 participants agreed to take part in the study. A total of 122 (84 per cent) were acceptable for analysis. The table below summarizes the socio-demographic characteristics of the sample.

Characteristic		**Per cent (n = 122)**
Gender	male	42.6
	female	57.4
Age	under 25	3.3
	25 - 35	50.8
	36 - 45	29.5
	46 - 60	16.4
Work experience	less than 1 year	1.6
	1 - 5 years	14.8
	6 - 10 years	29.5
	11 - 20 years	32.8
	more than 20 years	21.3
Previous experience with performance evaluation	yes	73.8
	no	26.2
Previous experience with performance targets	yes	65.6
	no	34.4
Previous experience with bonus payments	yes	50.8
	no	49.2

Table 45: socio-demographic characteristics of respondents

Considering gender, the ratio is rather balanced, with a slight majority of participants being female. More than half of the participants were younger than 36 years and nearly 70 per cent had work experience of more than 5 years. 21.3 per cent even had work experience of more than 20 years. Thus, the sample base is quite representative for employees working in the German service industry in 2012 in terms of gender, age and work experience (German Federal Statistical Office, 2015). Further, the majority of respondents had previous experience considering the three dimensions of the proposed

approach. Nearly two thirds had previous experience with performance evaluation and even half of the respondents had experience with bonus payments.

8.2.1.3 Data Collection

Since the respondents attended lectures and seminars at Hochschule Harz, the survey was conducted after these courses. At the beginning of the sessions, the respondents were provided with the participant information sheet to inform them that their participation is voluntary and that the survey is conducted anonymously so that the answers cannot be related to individuals.

Initially, the respondents were introduced to the aims of the research and to the proposed approach, which was illustrated using the same presentation that was used for the Focus Group Interviews. Subsequently, the same exemplary performance evaluation sheet that was used during the Focus Group interviews was explained and analysed. Eventually, respondents were given the opportunity to ask questions. Next, they were provided with the questionnaire, which was completed on-site and collected immediately after completion.

8.2.2 Data Analysis and results

The results of the survey were analysed on a descriptive level first. To gain more in-depth knowledge, an explanatory analysis was carried out subsequently. In the following sections, essential findings from both descriptive and explanatory analysis are being discussed. The overall results of the descriptive analysis are provided in annex 9. The overall results of the explanatory analysis are provided in annex 11 respectively.

8.2.2.1 Descriptive Analysis

To test for the questionnaire´s internal consistency, Cronbach´s alpha was calculated (Saunders et al., 2012). Since the questionnaire addresses several distinct topics, it did not seem appropriate to report alpha for the whole questionnaire. Therefore, alpha was calculated for each section (except the biographical section). The table below holds the results for Cronbach´s alpha, mean and standard deviation.

	Cronbach´s alpha	Mean	SD
Introductory section	0.443	1.710	0.590
Section A: general requirements	0.940	2.078	0.104
Section B: administrative and developmental purpose	0.752	1.844	0.109
Section C: Comparison of methods	0.726	1.684	0.581
Section D: Overall assessment	0.859	2.164	0.300

Table 46: Cronbach´s alpha, Mean and SD for the questionnaire´s sections

Cronbach´s alpha determines the internal consistency of items in dichotomous and/or multi-point formatted questionnaires. The alpha coefficient ranges from 0.0 to 1.0 indicating the inter-relatedness between items and the homogeneity of the construct. Thus, the higher the coefficient, the more reliable the generated construct is. There are different reports about the acceptable values of alpha, ranging from 0.70 to 0.95 (Tavakol and Dennick, 2011). The limit of 0.7 has been suggested as lower bound by several researchers, for instance Nunally (1978) or Hair et al. (2009) and is widely accepted in research (C. E. Pettijohn et al., 2001; Sudin, 2011). The reliability analysis of the questionnaire´s sections produced an alpha value above 0.7 for all sections except for the introductory part.

Since the introductory part evaluated the importance of several distinct characteristics, it is no surprise that the items of this section are of rather heterogeneous nature. In addition, this section does not deal with the assessment of the proposed approach itself. To identify the dimensions of the introductory part, a factor analysis was performed subsequently. It extracted four components that explain about 70 per cent of the total variance.

	Component			
	1	2	3	4
The PE should be fair		.755		
The Measures should be known	.772			
The Measures should be determined in cooperation with employees	.607			.452
The Method should be objective			.545	.595
The Measurement should not only focus on outputs				.895
The PE should identify weaknesses and potential for improvement		.768		
The whole PE process should be comprehensible	.753			
The results should be used as a basis for calculation bonuses or variable compensation			.855	

Table 47: Factor Analysis (pattern matrix) of introductory part

The components are related to comprehensibility of measures and the whole process (component 1), developmental purpose (component 2), administrative purpose (component 3) and the characteristics of measures (component 4). The further sections of the questionnaire relate to all of these issues.

In terms of assessing the importance of eight of the most cited requirements to PES (see chapter 4.3), the respondents evaluated “perceived fairness” as the most important characteristic by far (91.8 per cent). “Objectivity” (63.9 per cent) and the “ability to identify strengths and weaknesses” (62.3 per cent) were also considered to be “very important” by more than 60 per cent of the respondents. Contrary to previous studies, respondents did not find it very important that measures are determined in cooperation with the employees. Only 6.6 per cent assessed this characteristic as “very important”. 16.4 per cent even assessed it as “unimportant”. To use the results of a PES to calculate bonus payments was also not considered as “very important” by the majority of respondents. The table below summarizes the frequencies for the introductory part.

	Very important	Important	Neither important nor unimportant	Un-important	Very unimportant
The PE should be fair	91.8	8.2	-	-	-
The Measures should be known	44.3	37.7	13.1	3.3	-
The Measures should be determined in cooperation with employees	6.6	32.8	32.8	16.4	-
The Method should be objective	63.9	36.1	-	-	-
The Measurement should not only focus on outputs	55.7	32.8	6.6	3.3	-
The PE should identify weaknesses and potential for improvement	62.3	37.7	-	-	-
The whole PE process should be comprehensible	60.7	36.1	1.6	1.6	-
The results should be used as a basis for calculation bonuses or variable compensation	14.8	57.4	26.2	-	-

Table 48: Introductory Questions (questions 1 to 8) - Frequency distribution in per cent

By analysing the results for part A (general requirements for measuring performance) which are displayed in the table below, it is found that the proposed approach was evaluated very positive concerning all listed characteristics with approval rates[11] ranging from 88.5 per cent ("objectivity") to 78.7 per cent ("comprehensibility"). The characteristics "fairness" and "traceability of results" were actually rated as "very important" by more than 30 per cent (31.1 per cent and 32.8 per cent). Thus, the proposed approach was perceived particularly well regarding the characteristics "fairness" and "objectivity" that were considered most important by the respondents.

[11] The term "approval rate" is defined by adding up the values of the first two categories on the Likert Scale generally labelled as "strongly agree" and "agree"

	Strongly agree	Agree	Neither agree nor disagree	disagree	Strongly disagree
The results and statements of the provided result tables are comprehensible	21.3	62.3	9.8	6.6	-
The measures used for measuring performance are comprehensible	18.0	60.7	14.8	6.6	-
The results of the provided result tables are traceable	32.8	52.5	1.6	13.1	-
The PE is fair as a basis for measuring performance	31.1	49.2	14.8	4.9	-
The PE is an objective basis for measuring performance	16.4	72.1	8.2	3.3	-

Table 49: Part A – general requirements (questions A 1.1. to A 1.5.: measuring performance) - Frequency distribution in per cent

The assessment of the dimensions "determining performance targets" and "calculating bonus payments" produced similar results (see annex 9). However, the approach´s ability to meet the requirements considering the calculation of bonus payments was rated marginally lower across all characteristics. This may be due to the fact that the ability to calculate bonuses was not of high importance to the respondents. Especially the characteristic "comprehensibility" scored rather low (with 14.8 per cent of respondents found it "quite incomprehensible") considering the ability to calculate bonuses. One reason for this could be that – unlike the other two perspectives – the results of this perspective do not stand on their own, but are intermediates to calculate the bonus payments[12].

Part B dealt with assessing the method´s suitability regarding developmental and administrative purposes. Again, all listed characteristics achieved high approval rates ranging from 78.7 per cent (target setting) to 93.4 per cent (identifying individual strengths). Interestingly, the question whether the proposed approach can help to recognize where an employee is better than others is assessed rather ambiguous. While

[12] Alternative 1 was used for calculating bonus payments (see chapter 7.3.1.4)

this characteristic received the highest percentage from all characteristics listed in part B for “strongly agree” (39.3 per cent), it also received the highest percentage for “disagree” (6.6 per cent). This somewhat reflects the controversial discussions during the Focus Group interviews regarding the identification and naming of peers.

	Strongly agree	Agree	Neither agree nor disagree	disagree	Strongly disagree
The PE can help to set ambitious, but achievable targets	27.9	50.8	18.0	3.3	-
The PE can help to identify individual weaknesses and to improve performance	31.1	55.7	9.8	3.3	-
The PE can help to identify individual strengths and to develop them further	37.7	55.7	4.9	1.6	-
The PE can help to recognize where an employee is better than others	39.3	42.6	11.5	6.6	-

Table 50: Part B –administrative and developmental purpose (questions B 2.1. to B 2.4.) - Frequency distribution in per cent

The comparison of approaches in part C revealed some interesting insights. Compared to rather one dimensional approaches like simple rankings, written appraisals or appraisal interviews, the proposed approach is significantly perceived as “better”. Interestingly, these methods are the ones that were currently employed by the banks that participated in the Focus Group Interviews. The 360-degree feedback is of similar complexity to the proposed approach. However, only 18.0 per cent of respondents perceived the proposed approach as “better”, 19.7 per cent assed it as “about the same”. However, 39.3 per cent of respondents did not know the method at all. which makes this particular evaluation less generalizable. The following table summarizes the results.

Performance measurement: How does the proposed approach compare to:	Better	Worse	About the same	I do not know the method/ procedure
Simple rankings	80.3	6.6	8.2	4.9
Written appraisals by management	63.9	13.1	18.0	3.3
Appraisal interviews	59.0	11.5	24.6	3.3
360-degree feedback	18.0	3.3	19.7	39.3
Any other method you know	4.9	4.9	1.6	

Table 51: Part C – Comparison to other methods (question C 3.1.: measuring performance) - Frequency distribution in per cent

The comparison to other methods for determining targets and calculating bonuses came out similar. The vast majority of respondents rated the proposed approach as "better" than the listed alternatives. Even when adding other methods on their own, respondents rated the proposed approach usually as "better".

In final section D, respondents were asked for an overall assessment. Again, the prosed approach received high approval rates. For the purpose of evaluating employee performance, 36.1 per cent found the method to be "very suitable" and 52.5 percent assessed it as "suitable". For the other two perspectives the results were similar, but again the method´s suitability regarding the calculation of bonus payments was rated slightly lower. A small percentage of respondents (4.9 to 8.2 per cent) found the method "unsuitable" for the dimensions mentioned. No one evaluated it as "very unsuitable".

	Very suitable	Suitable	Neither suitable nor un-suitable	Un-suitable	Very un-suitable
For evaluating employee performance	36.1	52.5	6.6	4.9	-
For determining performance targets	32.8	55.7	6.6	4.9	-
For calculating bonus payments	29.5	49.2	13.1	8.2	-

Table 52: Part D – Overall assessment (question D 4.1.: general suitability) - Frequency distribution in per cent

Subsequently, the respondents were asked whether they would recommend the PES´ implementation in their organization. As a result, about one fifth (21.3 to 23 per cent) of the respondents would have recommended its implementation without supplements or changes. The majority of respondents (44.3 to 37.7 per cent) would have preferred to complement the proposed approach by other methods. Another fifth (21.3 to 16.4 per cent) would have implemented it equally to other methods. The participants were given the opportunity to add comments on the questionnaire. The most frequent comment considering the supplement of other methods was the wish for an additional "subjective component" like a written or oral assessment by management. Although many respondents rated this method "worse" than the proposed approach (see question C3.1), they opted for it as a supplement. In addition, some respondents wished for an additional assessment of "soft factors" like social behavior or team spirit.

Whilst only 3.3 per cent would have preferred the proposed approach as a supplement to other methods for evaluating employee performance, 10.0 per cent would have preferred it as a supplement for calculating bonuses. In total, 6.6 per cent of would not have recommended the proposed approach for evaluating employee performance at all. Again, this rate is slightly higher for the purpose of calculating bonuses.

	Yes, absolutely	Yes, complemented by other methods	Yes, equally to other methods	Yes, as a supplement to other methods	no
For evaluating employee performance	21.3	44.3	21.3	3.3	6.6
For determining performance targets	21.3	42.6	19.7	6.6	9.8
For calculating bonus payments	23.0	37.7	16.4	10.0	11.7

Table 53: Part D – Overall assessment (question D 4.2.: recommendation) - Frequency distribution in per cent

8.2.2.2 Explanatory Analysis

In a more in-depth analysis it shall be investigated whether the variables assessed in sections A and B are related to the participant´s overall assessment of the PES. For this purpose, the following models were defined:

Model A: variables of part A (dimension: evaluating performance) * approach´s general suitability (for evaluating performance) [independent variables: A.1.1, A.2.1, A.3.1, A.4.1, A.5.1; dependent variable: D.1.1)

Model B: variables of part A (dimension: determining performance targets) * approach´s general suitability (for determining performance targets) [independent variables: A.1.2, A.2.2, A.3.2, A.4.2, A.5.2; dependent variable: D.1.2)

Model C: variables of part A (dimension: calculating bonuses) * approach´s general suitability (for calculating bonuses) [independent variables: A.1.3, A.2.3, A.3.3, A.4.3, A.5.3; dependent variable: D.1.3)

Model D: variables of part B (administrative and developmental purposes) * method´s general suitability (for evaluating performance) [independent variables: B1 to B4; dependent variable: D.1.1)

Since the scatterplots between the independent variables and the dependent variable suggest a linear relationship for all models, a multiple linear regression analysis is performed. To obtain valid results, the data needs to meet several assumptions (J. Cohen et al., 2013). Besides a linear relationship (assumption (1)), the most cited assumptions are:

(2) The dependent variable should be measured on a continuous scale (i.e., it is either an interval or ratio variable)

(3) No or little multicollinearity of variables (two or more independent variables that are highly correlated with each other)

(4) No auto-correlation of standardized residuals/ independence of observations (i.e., independence of residuals).

(5) Multivariate normality (the error between observed and predicted values (i.e., the residuals of the regression) should be normally distributed)

(6) Homoscedasticity (the variances of standardized residuals are similar).

According to assumption 2, a multiple regression analysis requires continuous data. Since the questionnaire uses 5-point- Likert-Scales as response format, the obtained data are ordinal in character. Concerning the adequate use of statistic methods, there is an ongoing debate on how to analyze Likert type data appropriately. The debate revolves around two major views. On the one hand, some researchers support the “ordinal” view, claiming that Likert-Scales produce rank order data and therefore must be analyzed using non-parametric statistics (which are considered to be less sensitive and less powerful than parametric statistics). An advocacy of this view was put forward by

Jamieson (2004). On the other hand, there is also support and strong empirical evidence of the “intervalist” view. Supporters of this view argue that there is good reason to treat Likert-Scale data as interval data. They suggest that Likert-Scales (a collection of items) as opposed to individual Likert-Items are in fact interval data and may be analyzed parametrically. Carifio and Perla (2007) list a variety of studies that have shown that Likert response format produced empirically interval data. This was also confirmed by Lubke and Muthen (2004), who add the constraint that the data need to stem from a homogenous population. In this respect, it seems justifiable to use Likert-Scale data for a multiple regression analysis. In terms of a cautious approach, a stricter alpha level of .01 (instead of .05) was chosen for the analysis

Assumption 3 postulates that the independent variables must not correlate highly with each other. To detect for multicollinearity the Tolerance (T) and VIF values for each regression models were checked. Under the requirement $T < 0.01$ and $VIF > 10$, the models show no multicollinearity. To detect auto-correlation, the values of the Durbin-Watson test were investigated. With all d-values within a 1.5 to 1.9 range, it can be assumed that there is no auto-correlation in the multiple linear regression data. To check for multivariate normality (assumption 5) requires an analysis observing if the error between observed and predicted values (i.e., the residuals of the regression) is normally distributed. This was done by plotting the standardized residual values on a histogram with a fitted normal curve. In models A, C and D the values are approximately normally distributed. In model B the distribution is rather skewed to the right. For checking homoscedasticity, the standardized residuals were plotted against the unstandardized predicted values reviewing the Q-Q-Plots for each model. Since variances along the line of the best fit remain similar, homoscedasticity can be assumed (J. Cohen et al., 2013). The complete analysis is provided in annex 10. Since none of the assumptions is violated, the data was analysed using multiple regression.

The table below holds the results investigating the relationship between the approach´s suitability to meet general requirements to a PES and its perceived overall suitability for measuring performance (model A). With $p < .000$ the model is significant. Also the adjusted R^2 coefficient is relatively high (.469), indicating that 46 per cent of the variance in overall suitability may be attributed to the combination of variables in part A. The results indicate that the variables “comprehensibility” and “perceived fairness” are significantly related to overall suitability for measuring employee performance ($p < .01$). Furthermore, the standardized coefficient (beta value) indicates, that “perceived fairness” has the greatest statistical effect (.416) on overall suitability.

Variable	Mean	SD	Regression Coefficient	Standardized Coefficient (Beta)	Significance
The results and statements of the provided result tables are comprehensible	2.02	.760	.131	.130	.238
The measures used for measuring performance are comprehensible	2.10	.765	.361	.360	.000
The results of the provided result tables are traceable	1.95	.935	.-109	-.132	.160
The PA is fair as a basis for measuring performance	1.93	.810	.394	.416	.000
The PA is an objective basis for measuring performance	1.98	.616	.006	.005	.955
R^2 = .491; adjusted R^2 = .469; p value = .000					

*Table 54: regression analysis model A (methods ability to meet requirements for measuring performance * overall suitability for measuring performance)*

The adjusted R^2 for determining performance targets (model B) and calculating bonuses (model C) is .641 and .657 respectively (more detailed results are provided in annex 11). Both models are significant at 0.01 levels. Thus, more than 65 per cent of the variance of overall suitability for determining performance targets and calculating bonuses can be attributed to the identified variables. Again, “perceived fairness” and “comprehensibility” prove to be significant ($p < .01$). For determining performance targets, “comprehensibility” has the greatest impact (Beta = .526), for calculating bonuses “perceived fairness” has the highest standardized coefficient (Beta = .301).

Table 56 holds the results analysing the relationship between the approach´s suitability to account for administrative and developmental purposes and its perceived overall suitability (model D). The results show that adjusted R^2 is lower than in the previous models, accounting for 28 per cent of the total variance. The only variable that is significantly related to the overall suitability for measuring performance is “ability to determine ambiguous, but achievable targets” with an impact of Beta = .309. Thus, the only variable that has a modest impact is of administrative nature. Variables reflecting developmental purposes are not significantly related to overall suitability.

Variable	Mean	SD	Regression Coefficient	Standardized Coefficient (Beta)	Significance
The PA can help to set ambiguous, but achievable targets	1.97	.771	.309	.331	.002
The PA can help to identify individual weaknesses and to improve performance	1.85	.724	.185	.174	.118
The PA can help to identify individual strengths and to further develop them	1.70	.639	.142	.118	.239
The PA can help to recognize where a participant is better than others	1.85	.869	.072	.082	.349
R^2 = .302; adjusted R^2 = .278; p value = .000					

*Table 55: regression analysis model D (administrative and developmental purposes * overall suitability for measuring performance)*

The descriptive analysis already indicated that the methods suitability to calculate bonus payments received uniformly lower ratings than the other two dimensions (measuring performance and determining performance targets). This may be due to the fact that employees did not perceive this particular feature as very important for a PES. To investigate this hypothesis, a regression analysis exploring the relationship between the perceived importance of calculating bonuses (I.7) and the method´s perceived suitability (D.1.3) to do so was performed. The alpha level is .01.

Variable	Mean	SD	Regression Coefficient	Standardized Coefficient (Beta)	Significance
The results should be used as a basis for calculating bonus payments	2.12	.638	.294	.216	.018
Dependent variable: Method´s general suitability for the purpose of calculating bonus payments					

*Table 56: regression analysis importance of the feature "calculating bonuses * overall suitability for the purpose of calculating bonuses*

The results show a significant relationship between these variables ($p < .05$). Thus, the employees´ perception considering the importance of calculating bonuses had an impact on their assessment of the proposed approach. This influence however is of rather moderate manifestation (Beta = .216).

A second analysis investigated the relationship between the respondent´s overall assessment of the method and their biographical background. The analysis included three different types of data levels:

- Overall suitability/ recommendation: ordinal data (commonly treated as interval data)
- Age, work experience: ordinal data
- Gender, previous experience with performance evaluation: nominal data

Thus, the chi-square test was selected to investigate the association of the variables, since this test also works with nominal data. The Chi square test it is also sensitive to the distribution within the cells (if they have fewer than 5 cases). This issue was addressed by additionally running a Fisher´s exact test that is more appropriate for small cell counts (Myers et al., 2010). As a measure of association, the chi-square test calculates the Phi coefficient which is equivalent to the correlation coefficient *r*. However, Phi mainly is applied for two-by-two tables since it has no upper limit to its value for larger tables. An extension of this approach is Cramer´s V, which can be calculated for larger tables (Kuckartz et al., 2010). For the analysis a confidence level of alpha = .05 was determined.

dependent variable: General suitability of the approach (for measuring employee performance)				
independent variable	Phi	Cramer´s V	p-value (sign.)	p-value (Fishers exact)
Age	.538	.310	.000	.000
Experience with Performance Evaluation	.409	.409	.000	.000
Gender	.388	.388	.000	.000
Work experience	.435	.251	.027	.049

*Table 57: Chi square analysis biographical data * overall suitability for measuring performance*

Considering the overall assessment, all biographical variables were found to be significantly related to the method´s perceived suitability for measuring performance. With Cramer´s V ranging from .251 (work experience) to .409 (previous experience with performance evaluation) the correlation is rather moderate. However, the employees´ demographic and biographic background seems to influence their perception of the method.

For a more comprehensive analysis, the descriptive data was included. The table below shows the assessment of the approach´s overall suitability differentiated in terms of "age". Apparently, younger respondents (up to 35 years) tend to find the proposed approach more suitable than older respondents.

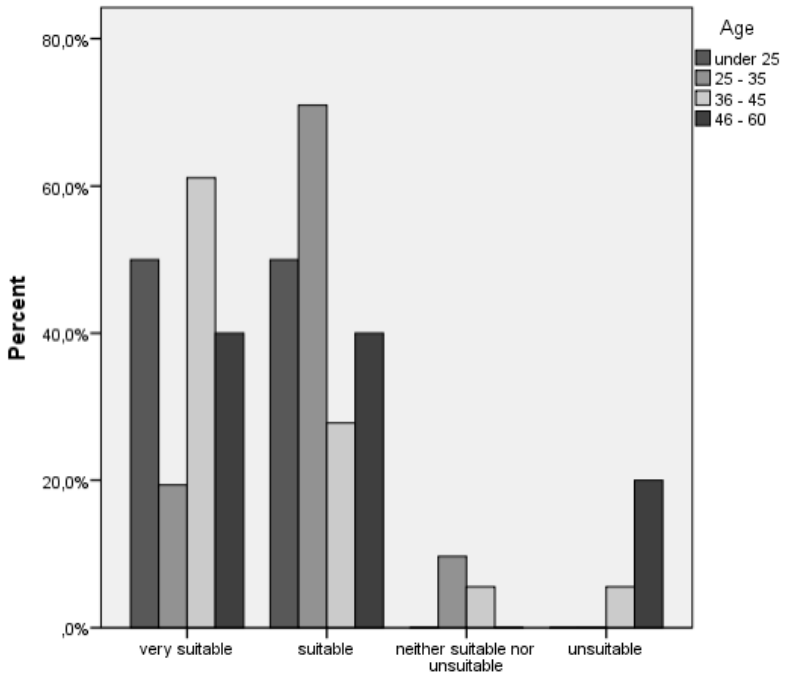


Figure 38: overall assessment in terms of to age

This argumentation is supported by analysing the assessment in terms of "work experience". Employees with less work experience (less than five years), who generally tend to be younger, evaluated the method more positive than employees with more work experience.

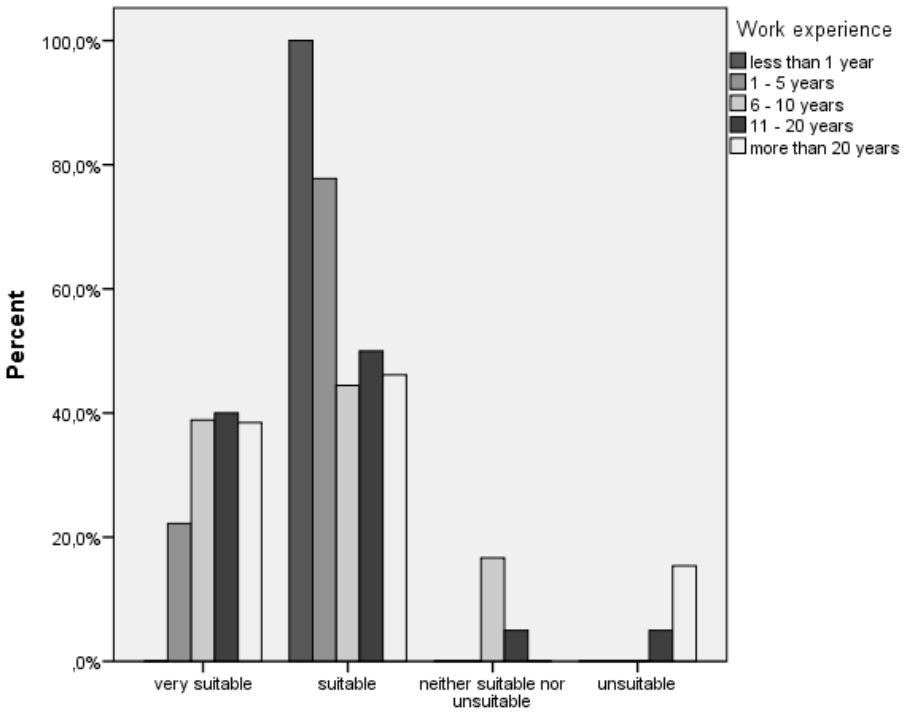


Figure 39: overall assessment in terms of work experience

The variable "previous experience with performance appraisals" had the greatest statistical impact on the overall assessment. Apparently, employees who already experienced a performance evaluation perceived the proposed approach as more positive than the ones who lack this experience. Thus, none of those participants evaluated the approach as "unsuitable".

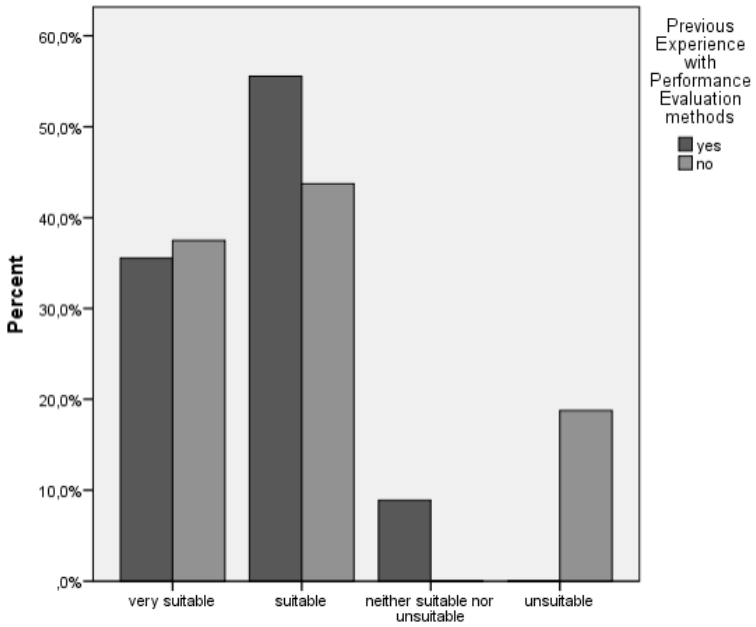


Figure 40: overall assessment in terms of previous experience with performance appraisal

Regarding the assessment in terms of “gender”, the frequency charts indicate that none of the male respondents found the proposed approach unsuitable. However, more female respondents than male respondents assessed it as “very suitable”.

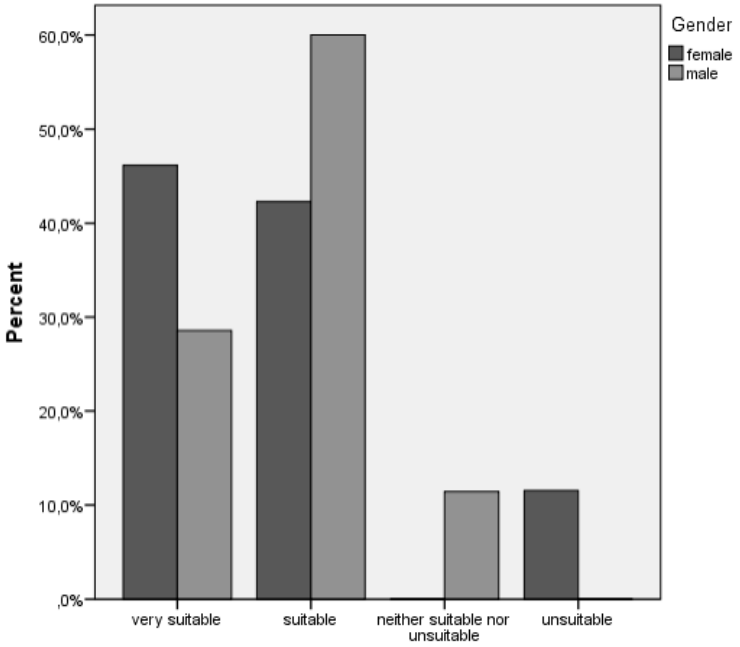


Figure 41: overall assessment in terms of gender

Considering the approach´s suitability for determining performance targets respectively for calculating bonuses, only “age” was found to be significantly related to the overall assessment, with a Cramer´s V of .318 (targets) and .320 (bonuses).

8.3 Findings

The findings from the method´s assessment by all major stakeholders are reported in the following sections. First, the strengths and weaknesses of the proposed method that were identified are highlighted. Further, issues that were discussed or evaluated controversially are illustrated.

8.3.1 Strengths

One major point of agreement during the Focus Group Interviews was the method´s superiority to other performance evaluation methods. Nearly all participants pointed out that the proposed method provided richer and more comprehensive results than the methods they currently apply. This was also confirmed by the assessment by employees, who rated the proposed methods as superior to other evaluation methods they know. Further, the method was perceived as very fair by all stakeholders. This feature has been

assessed as “very important” or “important” by 91.8 per cent of the questionnaire´s participants and was found to have a significant impact on the method´s overall perception by employees. Regarding the importance of PES characteristics, “perceived fairness” was followed by “objectivity” (63.9 per cent) and the “ability to identify strengths and weaknesses” (62.3 per cent) in employees´ assessment. The proposed approach received particular high ratings considering these three characteristics.

Focus group´s participants evaluated the results as very helpful for counselling employees considering training needs and career goals. They valued the fact that the method provides more in-depth information than a simple rating or an assessment by line managers. In this respect, the performance evaluation sheet was assessed as a very concise and helpful tool. Although both managers and workers´ council stated, that they found the results comprehensible, they raised the questions whether employees would feel the same. The questionnaire study showed that comprehensibility not only received high approval rates (83.6 per cent) but also – along with perceived fairness - had a significant impact on the method´s overall perception.

Furthermore, there was major approval considering the calculation of bonus payments by all stakeholders. In some Focus Groups there were discussions whether to impose limits for bonus payments or whether to calculate bonus payments for managers in the same manner. In terms of strategic purposes, the method was perceived as very suitable for translating business strategy into individual goals by the managers´ groups. In this respect, dividing the whole performance process into several performance stages was perceived as very helpful and evaluated as “a good way to review performance from different angles”. Further, it was appreciated that, in order to define appropriate measures, management was forced to clearly define, prioritize and communicate their organizational goals.

8.3.2 Weaknesses and Limitations

The assessment, in particular the Focus Group interviews, also identified some weaknesses and limitations for applying DEA to evaluate employee performance. First, stable DEA results require a large number of DMUs, especially when a variety of inputs and outputs is considered. As a rule of thumb, the total number of input and output variables should not exceed one-third of the total number of DMUs (R. R. Thomas et al., 1998; Wagner et al., 2003). Thus, DEA may not be appropriate for smaller organizations or very heterogeneous groups. Second, given the variety of input and output variables, DEA relies on the availability of valid data. For the case study, most of the data were

made available by the bank´s controlling department. If the data is not readily available, data collection may cause a disproportional effort for some banks.

A third issue, that was pointed out by both managers and workers´ council, was the method´s transferability to other job profiles. In the context of this case study, this was especially true for employees working in the bank´s back office whose performance process often is more diverse. Thus, groups of employees with a similar performance process may be smaller than in the front office section. In addition, whilst their inputs may be similar to front office employees, the performance output of employees working in the back office is more difficult to capture. Thus, to examine the transferability and applicability of the DEA-based PES to back office employees is an important task for future research.

Another issues that was raised by all stakeholders was, that performance targets calculated by DEA were perceived as too demanding in some cases. This refers to a concern that often is discussed in DEA literature (N. Adler et al., 2002; Wang and Chin, 2010). . Since DEA calculates its target values by linear optimization technique, it assumes that the calculated targets are achievable by reducing (some) input factors or by increasing some output factors. In reality, this may not always be possible since there may be additional factors to target achievement (e.g. personality or motivation) that were not included in the calculation but have an impact on performance. Therefore, the calculated targets should rather be regarded as maximum or upper limit.

8.3.3 Controversial issues

Although there mostly was agreement across all groups of stakeholders considering the method´s strengths and weaknesses. some issues were assessed rather controversial.

Among the most divisive issues was the method´s objectivity. Whilst many participants of both Focus Group and questionnaire assessment praised the method´s objectivity, some claimed that “subjective factors” including a supervisor´s personal impression of an employee should also be considered in order to gain a comprehensive assessment. Therefore, some participants suggested complementing the method by a rather subjective component. This was stated by participants from managers´ groups and also from participants of the questionnaire study. It should be noted that the call for a subjective component was not supported by the groups from the workers´ council. Further, the assignment of weights was also assessed rather ambivalent in particular in the Focus Groups. Thus, some participants supported the idea of imposing weights to stress the importance of some measure whilst others claimed that this would contradict

the method´s perceived fairness which they highlighted as one of its “greatest assets”. Again, the members of the workers´ council evaluated the method´s independency as very beneficial.

Eventually, the identification and naming of peers was met with varied receptions. Managers´ groups in general were rather fond of the idea since the existence of a real-life peer showed that improvement in some areas can be done. Also, the idea of teaming up employees with complementary strengths and weaknesses was evaluated as beneficial. However, some of the participants in those groups pointed out that naming peers could cause a negative working atmosphere. The members of the workers´ council were even more cautious considering this matter. They agreed that it should be communicated to the employees that there is a peer. However, they rejected the idea of naming peers not only because they feared a negative impact on motivation but also due to confidential reasons. Also, some employees addressed this issue, noting that peers should not be named.

9 Conclusion

The concluding chapter of the thesis outlines the major findings and associated contributions of this research. It will also discuss how this research is linked to previous research findings and the implications it may have on theoretical and practical level. Finally, this chapter points out both limitations and likely directions for further research activities.

9.1 Research questions and major findings

The aim of this thesis was to investigate DEA´s suitability to serve as an evaluation method for employees´ performance in the service industry. Thus, after analysing and reflecting previous research on all relevant areas, DEA was applied for evaluating the performance of account managers in a German Cooperative bank using a case-study approach. The research objectives aimed at assessing DEA´s suitability on technical and organizational level. In the following sections, the major findings considering the research objectives and their subsidiary research questions (A1 to A5 and B1 to B3), which were postulated in chapter 1.4.1 are outlined.

9.1.1 DEA´s technical suitability for evaluating employees´ performance in the service sector

Since DEA was applied in the service industry, the question whether evaluating service performance actually places advanced requirements to the underlying evaluation method (research question A1) needed to be investigated. From the literature review on service research (including Vargo and Lusch, 2004, 2008; Lovelock and Gummesson, 2004; Grönroos and Voima, 2013, Grönroos and Ojasalo, 2004, Schmenner, 2004) it could be concluded that there is no “unique” service, but different service types that relate to the IHIP characteristics in varying degrees considering labour intensity, interaction and customization (Lovelock and Gummesson, 2004; Vargo and Lusch, 2004, 2008; Grönroos and Voima, 2013). Both Schmenner (1986, 2004) and Silvestro (1999) offer classification schemes that identify three distinct service types: “professional services”, “service shops” and “mass services”. Applied to banking services, the classification shows that banking includes all three types of services, which places different requirements to management (Schmenner, 2004). In particular, services falling into the categories of “professional services” and “service shops” offer a high degree of interaction and customer integration and therefore strongly rely on employees´

performance (Benkenstein et al., 2017). However, profitability of these services and performance of employees producing them is difficult to ascertain due to the variety of input factors, such as the number of customers or employees´ experience, and output factors, for example the number of customers counselled or products sold (Grönroos and Ojasalo, 2004). Further, to produce a service that meets the customers´ demands for quality, employees may have varying production processes (Grönroos and Ojasalo, 2004; Grönroos and Voima, 2013) applying different input and output combinations to produce a service. Therefore, it could be concluded that an adequate performance evaluation method for service performance needs to consider all these peculiarities and should meet requirements that go beyond traditional performance evaluation.

Investigating what advanced requirements are important for evaluating service performance (question A2), literature review on Performance Management (including Chartered Institute of Personnel and Development, 2016a; Armstrong, 2018, 2015; Shields, 2015; Jackson et al., 2010) showed that an adequate evaluation method requires taking several input and output factors into account simultaneously (Shields, 2015; Jackson et al., 2010). In addition, it should be able to discriminate between different production processes and to give detailed information about relative distance between employees (Vargo and Lusch, 2008). Also, it should account for factors that are beyond the employees´ control but nonetheless affect their performance (Shields, 2015), as the number of inhabitants in a catchment area, for instance. In terms of general requirements to a performance evaluation, the analysis of a variety of studies (Armstrong, 2018; Chartered Institute of Personnel and Development, 2016a, Sudin, 2011) identified several frequently cited recommendations to be of high importance. In this respect, it was found that “perceived fairness” and providing detailed feedback information were the most cited features for conducting a proper performance evaluation (Chartered Institute of Personnel and Development, 2016a; Sudin, 2011).

Considering the pros and cons of traditional approaches for performance evaluation (question A3), the comprehensive analysis on performance evaluation methods in chapter 3.2.6 (Shields, 2015; Armstrong and Taylor, 2014) revealed several shortcomings considering both general and advanced requirements. Thus, traditional absolute approaches fall short considering perceived fairness and feedback information since they are often perceived as biased and are often not able to provide in-depth improvement information (Armstrong and Taylor, 2014). Comparative approaches, on the other hand, fail to integrate and aggregate several input and output factors. Also, most approaches fail to discriminate between different production processes or to account for uncontrollable factors (Shields, 2015).

The review of DEA literature (including Cook et al., 2014; Cooper et al., 2011; Banker and Morey, 1986a; Warning, 2014) and the analysis of the underlying methodology (Charnes et al., 1978) provided convincing evidence that DEA may be suitable to address the shortcomings of traditional approaches (question A4). Due to its non-parametric nature, DEA requires no a priori specifications of the parametric form of the production correspondences (Banker and Morey, 1986a). Also, it includes multiple inputs and output measures simultaneously, being able to account for uncontrollable factors (Cook et al., 2014). Moreover, DEA identifies employees that operate with similar input-output combinations (Cooper et al., 2011). Despite its application in various fields since first proposed in 1978, there is only very little experience on applying DEA on individual level (Warning, 2014).

Addressing research question A5, DEA´s application to evaluate employee performance, which was illustrated throughout chapter 7, provided equitable evidence that it gives reasonable results on individual level, identifying good and weak performers, indicating areas and levels of possible improvement and indicating peers as best-practice benchmarks. Considering the robustness and applicability of results, the subsequent sensitivity analysis has proven to be of crucial importance within the evaluation procedure. To reduce complexity most of the information was presented in a user-friendly dashboard-style. Further, all individual results were aggregated into performance portfolios to inform HR decisions on organizational level. The subsequent assessment of the method by all major stakeholders showed that the customized presentation and aggregation of the extensive results was important for the comprehensibility of results and the perception of the whole method.

9.1.2 DEA´s organizational suitability for evaluating employees´ performance in the service sector

Although Performance Management (PM) is a key function of HRM which organizations need to attend to irrespective of their specific HR strategic focus (Albrecht et al., 2015), there is a growing body of research questioning HRM and PM activities to be mutually beneficial (e.g. (Godard, 2004; Kramar, 2014). Thus, some HRM scholars note an absence of employees´ perceptions and claim to rather focus on incorporating the views and experiences of the wider workforce (Keeble-Ramsay and Armitage, 2015; Kramar, 2014). This thesis aimed at addressing this lack of information. Thus, DEA´s organizational suitability was assessed including the views of all major stakeholders. Considering employees´ performance evaluation, the CIPD (2016a) pointed out that what is important is not so much the procedure itself but rather the reactions to it.

Addressing research question B1 by investigating the general perception of DEA´s procedure and results, the assessment showed that the proposed method was evaluated very well across all groups of stakeholders. It was found that DEA was evaluated particularly well regarding the characteristics “fairness” and “objectivity”, Also, both characteristics were found to be significantly related to the employee´s overall perception of the approach. Interestingly, employees, who already experienced performance evaluation, rated the proposed approach significantly better than employees who lacked this experience. Another noteworthy finding is, that all stakeholders perceived the proposed method as “better” compared to traditional methods including simple rankings, written appraisals or appraisal interviews. Further, all groups agreed that the results were way more extensive and detailed than the results provided by other methods they knew. These findings support the assumption that DEA is able to address some of the shortcomings of traditional evaluation methods.

Considering DEA´s suitability to provide a basis for both administrative and developmental purposes (question B2), all stakeholders assessed the method as being appropriate. Managers in particular valued the results for supporting decisions about promotions or transfers. Regarding developmental purposes, a major requirement to any performance evaluation is its ability to provide clear feedback information and direction for improvement (Becker et al., 2011; Latham et al., 1993). Moreover, the method should enable an equitable treatment of different groups of employees. Again, “perceived fairness” and “reduction of bias” were found to be among the most important issues for the acceptance of feedback (Chartered Institute of Personnel and Development, 2016a). In line with these findings, the assessment showed that the provided feedback information was perceived as very helpful to identify areas of improvement and training needs across all groups. Employees in particular, valued method´s ability to identify individual strengths. This characteristic was also assessed as very beneficial by both managers and workers´ council, who also valued the method´s ability to incorporate uncontrollable factors. Moreover, they emphasized the benefit of not only rating employees´ performance, but also being able to point out areas employees need to improve and to identify circumstances under which they cannot improve. In line with previous studies including Smither et al. (2005), Aguinis (2012) and the CIPD (2016a) those findings are a strong case for applying rather strength-based approaches to obtain feedback information.

Considering the method´s suitability for calculating bonuses and performance targets (question B3), managers during the Focus Groups interviews, confirmed that, when it comes to combining employee performance and remuneration, methods are often not

properly designed. In this respect, managers and the workers´ council assessed the calculation of bonuses by DEA as being more fair and concise than traditional methods banks currently use. Employees´, however, rated the method´s ability to calculate bonus payments marginally lower than other characteristics. This is in coherence with the fact, that employees rated this particular characteristic not as highly important at all. Members of the workers´ council claimed that managers´ bonus payments need to be calculated in the same manner for managers in order to establish sustainable incentive structures. This claim was controversially discussed by managers, since they felt it was hard to define adequate measures, but they generally supported the claim. This finding supports the view of authors who criticized HR-practices for having contributed to the recent GFC (MacKenzie et al., 2012; Psychogois et al., 2016) or claimed a disconnection of HR managers, who were tied to financialised practices and measures and therefore lacked the capacity to enabling stable conditions and to engage in people-focused management (Thompson, 2011). Thus, an equal calculation of bonus payments based on long-term measures could be an important step to establish sustainable incentive structures.

The calculation of performance targets by DEA was in general assessed as fair and reasonable by all stakeholders. However, some targets were perceived as too demanding, which echoes an issue that is often discussed in DEA literature. Thus, the Focus Group discussions on this topic are consistent with previous studies (Adler et al., 2002; Wang and Chin, 2010b), stating it may not always be possible to fully achieve DEA´s targets since there may be additional factors to target achievement (e.g. personality or motivation) that were not included in the calculation but have an impact on performance. In the Focus Group discussions, some participants suggested to regard the calculated targets as maximum or upper limit.

The assessment also pointed out some drawbacks of DEA for evaluating employee performance. Considering its transferability to other job profiles, several managers found it hard to develop criteria for employee´s working in the back office. Although some measures were found that would work for back office employees, they mostly were not available or not concise. Second, DEA relies on the availability of valid data. As the case study showed, some data were not readily available (e.g. data about customer satisfaction) and data collection would have caused a disproportional effort. Third, stable DEA results require a larger number of DMUs, especially when a variety of inputs and outputs is considered. This may be an issue when evaluating smaller groups of employees.

9.2 Relationship to previous research

As outlined in chapter 3, the benefits and impact of PM on performance is still questioned by some HRM researchers. In particular, there has been some major criticism considering the role of HRM in general, including the question how HRM and HR practices may have contributed to the recent global financial crisis (GFC) (MacKenzie et al., 2012). However, there is a growing body of evidence that PM can be a valuable process if carried out properly (Chartered Institute of Personnel and Development, 2016a). The findings of this study are consistent with this previous research. On the one hand, they indicate that the proposed method provides performance information that is valuable for all major stakeholders participating in the evaluation procedure. On the other hand, participants in the assessment pointed out some issues that should be considered in order to do the evaluation "properly". Among those issues is the cautious selection of measures, since poorly defined or selected measures (e.g. short-termed measures, measures neglecting the views of major stakeholders) may lead to unintended behaviour and results. This finding seems to be in line with the arguments of Thompson (2011), Star et al. (2016) and Bititci et al. (2006), who claim that not only design, but also the execution and the organizational context, including management style, organizational culture or the country´s economic context, shape a performance evaluation and may lead even a well-designed PES to failure.

To understand under which specific conditions PM evaluations can actually deliver the promised results and how the whole procedure should be conducted properly, the work of some researchers reviewed in chapter 4 has identified a number of requirements. Most of this studies, including Cocca and Alberti (2010), Pettijohn and Pettijohn (2001) and Sudin (2011), characterized "perceived fairness" of the results and procedure as the most important requirement to a proper performance evaluation. The study´s findings seem to echo this assertion. Thus, the assessment showed that employees rated "fairness" as most important characteristic. Moreover, "perceived fairness" was found to have a significant impact on the employees´ overall perception of the proposed method. Also, the other stakeholders confirmed this impression.

Another requirement that is frequently cited to be of crucial importance is "objectivity" (Kondrasuk, 2012; Chartered Institute of Personnel and Development, 2016a). Again, the study´s findings supports this assertion at first sight. Taking a closer look on the results gives a more differentiated picture. Thus, despite all stakeholders emphasizing the importance of "objectivity", several participants from both managers´ and employees´ groups wished to add a rather "subjective component" to the evaluation. Moreover, unlike "perceived fairnesss", the method´s perceived objectivity was not found to be

significantly related to the method´s overall perception in the questionnaire study. Hence, these findings are partly at odds with previous findings and should therefore be subject to further research.

The literature review into the general aims of performance evaluation on individual level showed that organizations typically use the performance information for two main aims: administrative and developmental purposes (Kondrasuk, 2012; Meyer, 1991; Thomas and Bretz, 1994). Whilst traditionally the focus of PM was on accountability and on administrative purposes, current studies including Cappelli and Tavis (2016) or the CIPD (2016a) found that a developmental orientation and strength-based approaches became more popular during the last decade. The findings of this research are broadly in line with these notions. Although the case-study was carried out in the banking industry, which traditionally is rather focused on accountability (Federal Association of German Banks, 2017), all stakeholders in the assessment put their focus rather on developmental issues, for example DEA´s ability to identify weak and strong performers and to provide concise feedback information. Particularly managers and members of the workers´ council valued that DEA identifies groups of employees employing different input-output structures and therefore are not comparable to each other. This may provide some evidence to research, supporting Grönroos´ and Voima´s (2013) arguments that manufacturing based productivity models do not apply to (all types of) services due to heterogeneity in production. Moreover, this finding supports the claim of several HRM Scholars, including Boxall (2013), arguing that an organization´s labour force is not homogeneous and their skills and knowledge therefore are unique.

9.3 Implications

As a synthesis, the following section provides contributions or implications from this research that may impinge on existing theories or practical applications.

9.3.1 Theoretical Implications

This research was mainly supported by a comprehensive literature review drawing on distinct areas of research, including research on Data Envelopment Analysis (Charnes et al., 1978; Cook et al., 2014; Cooper at al., 2011; Paradi et al., 2011), Human Resources Management and Performance Management (Chartered Institute of Personnel and Development, 2016a; Shields, 2015; Kramar, 2014; Schläfke, 2012) and Service Management (Grönroos and Ojasalo, 2004; Vargo and Lusch, 2004, 2008; Gummesson and Lovelock, 2004). Also, considerations on Research Methodology (Yin,

2017, 2013; Saunders et al., 2012; Bryman, 2015; Johnson and Onwuegbuzie, 2004) informed this study. The following section reflects on where this research may have contributed to the theoretical discourse in some of these areas by providing some critical evidence or supporting specific arguments.

9.3.1.1 A pragmatic mixed-methods approach to evaluate a DEA application

To address the research objectives and questions, the author adopted a pragmatic research philosophy applying a mixed-methods approach in an iterative procedure. Despite its shortcomings, including a rather time-consuming data collection and analysis process (Johnson and Onwuegbuzie, 2004), the methodology has proven to be very useful to collect the needed data and to do a proper analysis. Since the application of DEA was investigated using an embedded single-case study approach, the author followed suggestions of Yin (2017) and Saunders et al. (2012) that case study research always should rely on multiple sources of evidence and by this should deliberately triangulate the evidence from these multiple sources to confirm and corroborate the findings. Thus, this research was based on a mix of qualitative and quantitative methods for data collection and analysis. Also, to include multiple sources and by this addressing concerns considering the validity and reliability of single-case findings, the assessment of DEA´s organizational suitability by the stakeholders was expanded to groups of managers and workers´ councils from other regional cooperative banks enhancing the case´s context. Since there are only few references in literature on single-case studies that included units outside the direct organizational context of the case, the present research may add some evidence to enhancing the context of single-case studies.

9.3.1.2 Controversy on imposing weights and importance of sensitivity analysis in DEA

Among DEA scholars there is an ongoing discussion on whether to assign weights manually to emphasize the importance of some input or output factors or to incorporate management´s view. In rejection of this claim, some authors including Pedraja-Chaparro et al. (1997) argue that the incorporation of weights deprives DEA of one of its most valued features and makes it less objective. Others, including Thanassoulis (1995) or Paradi and Schaffnit (2006) support the incorporation of weights for the purpose of expressing views of value for the individual inputs and outputs. Considering DEA´s application on employee level, this study showed that employees benefit from DEA´s unconstraint distribution of weights. Thus, when weights were imposed, the mean score

of the data set dropped slightly to significantly. Moreover, this study offers suggestive evidence that the debate over imposing weights that exists in research is also held in practice. Thus, managers in the Focus Group interviews discussed this matter equally controversial. Whilst many valued the fact that the proposed approach calculates weights by mathematical optimization techniques (and thus is independent from bias) others were in favour of the possibility of setting weight restrictions in order to stress the importance of some measures. Thus, the decision about the incorporation of weights will probably remain a case-by-case decision.

Considering the robustness of results, a subsequent sensitivity analysis - to identify variables or other DMUs that were of high influence - proved to be an important step within the evaluation procedure. In particular, since administrative and developmental decisions may be based on the results. In general, the results proved to be quite robust. However, in three performance stages outliers were identified. Since literature review brought up no evidence of standardized procedures for DEA´s sensitivity analysis, an iterative procedure including the correlation analysis of variables, the assignment of weights and a supereffciency analysis was developed. This procedure has been documented as process workflow and may allow its use and further development in future DEA studies, in particular if they are aimed on evaluating employees´ performance.

9.3.1.3 Integrating the organizational and the individual perspective

Since the term “performance management” relates to both, the performance of individuals and the performance of the whole organization, the vast majority of PM approaches focuses either organizational or on individual level (Yadav and Sagar, 2013). However, there is a growing cognizance among researchers and practitioners that there is no single PM process or approach that guarantees sustainable success and competitive advantage. Rather performance management on individual and organizational level should be interrelated in order to align corporate strategies with operational targets as well as organizational performance with individual development (Pun and White, 2005; Sebald and Jacob, 2015). In consequence, a holistic Performance Evaluation System should integrate the organizational and the individual perspective.

The assessment of the proposed method among managers showed, that the method´s suitability to translate business strategy into individual goals was among its most valued characteristics. They acknowledged that the method forced management to clearly define and prioritize their organizational goals. Moreover, a correlation analysis of performance scores showed that the distribution of calculated bonuses adequately

reflected the hierarchy of strategic targets that were previously defined by the bank´s management. This may also serve as an indicator for the method´s ability to implement and cascade strategic targets into individual performance targets.

Although DEA´s suitability for evaluating organizational performance was not subject to this research, some links for integrating DEA´s results on employee level into approaches on organizational level shall be pointed out. Most obviously, DEA could be integrated into an overall Business Analytics approach as proposed by Davenport (2006), which became also popular in the field of PM by supplementing traditional approaches with sophisticated and analytical decision-making tools (Schläfke et al., 2012). This would be particular advantageous since DEA is on the intersection of traditional approaches and analytical tools and therefore meets requirements from both sides. As a prerequisite, the procedure would need to be supported by a sophisticated IT-tool. From a more traditional focus, DEA could inform the internal perspectives of a BSC (or a comparative BSC as proposed by Kanji and Sà (2002)). Integrating DEA´s results into a BSC could advance existing approaches as proposed by Tan et al. (2017). Although their approach did not focus on individual level, they found that integrating DEA results into a BSC allowed an analysis of underperformance in four dimensions and helped to identify inter-relationships between the dimensions. In the form of an HR scorecard this concept could be transferred to the area of HRM.

9.3.1.4 Including the view of all stakeholders

Despite being essential for the organization in order to employ and develop people (Armstrong, 2018), the concept of HRM has recently been facing criticism for neglecting its intended purpose to serve both sides, employees and managers or shareholders (Keeble-Ramsay and Armitage, 2015). Thus, the stakeholder theory established by Freeman (1984), that fosters recognition of the fact that several stakeholders are voluntarily or involuntarily contributing to the performance and the success of the organization, has not received much attention from the area of HRM and PM yet (Freeman et al., 2010). In consequence, research into PM and evaluation of different PM approaches was mostly limited to managers´ perceptions whilst employees´ interests and attitudes often were neglected.

More recently, in the evolving field of sustainable HRM, researchers emphasize the importance of considering stakeholders' impact and wellbeing while still achieving financial outcomes for the organization (Kramar, 2014). Thus, stakeholder theory challenges the supremacy of one stakeholder´s interest. Taking up the call that research needs to focus on incorporating the views and experiences of all major stakeholders

(Kramar, 2014; Keeble-Ramsay and Armitage, 2015) this research included workers´ council and employees as equal groups besides managers into the method´s assessment.

The subsequent analysis of results showed that the three groups of stakeholders in general agreed over the strengths and weaknesses of the proposed method. However, it also showed that the groups sometimes put varying emphasis on characteristics. Workers´ council groups, for example, completely rejected the idea of assigning weights to some variables or to supplement the method with a more "subjective component", whilst the other stakeholders discussed this issue rather controversial. Employees, on the other hand, particularly valued the method´s fairness. In summary, all groups of stakeholders found it important that their opinion is heard and taken into account when designing a performance evaluation, which is a supportive argument to consider stakeholder theory when planning HR activities.

9.3.2 Practical Implications

The following sections presents implications that may be of practical relevance to the design and execution of a proper performance evaluation.

9.3.2.1 Choosing appropriate measures

To obtain meaningful results that reflect actual performance, measures must be chosen carefully. Micheli and Mari (2014) argue that since PM is mostly concerned with social subjects, whose performance is often complex and difficult to define and measure, the definition and collection of performance measures is a challenging task. This argument was confirmed by the assessment among the bank´s managers who provided the data for the case study. During the assessment, they pointed out that, although most measures were provided by the controlling department, the effort for data collection was perceived as quite high. Also, they perceived the prior process of defining adequate measures as very extensive.

To prevent selecting measures that were short-termed (Keeble-Ramsay and Armitage, 2015), lead to an increase in unethical behaviour (Welsh and Ordóñez, 2014) or were only selected because they were easily accessible (Micheli and Mari, 2014), the selected measures were reviewed using a performance measure sheet based on the recommendations of Neely et al. (1997). This procedure led to the removal of some

measures that did not meet the requirements and to the incorporation of others that partly were harder to acquire. Therefore, the case-study at hand appears to support Micheli´s and Mari´s (2014) rather pragmatic argument that a criterion for "adequate measurement" is a cautious trade-off between acceptable quality and available resources.

9.3.2.2 Presentation of results

Considering DEA´s results, some DEA scholars, including Paradi and Schaffnit (2004) point out that the results sometimes fail to impress management due to the fact that they are presented in a manner that is too complex and hard to understand for practitioners. To present the results in a more convenient manner for managers, workers´ council and employees, a performance evaluation sheet based on a dashboard-style was created for each employee. In the assessment, it was found that although managers feared that the results may not be comprehensible for employees, the questionnaire survey showed that comprehensibility was not an issue for employees. Participants across all groups pointed out, that the user-friendly presentation facilitated the comprehensibility of results. Another important issue considering the presentation of results was breaking down the performance process into distinct stages of performance. On the one hand, this also contributed to reducing complexity. Even more important, the consideration of several stages may have contributed to increase acceptance and motivation since most employees showed above-average performance in at least one stage.

9.3.2.3 Selection of an evaluation method and issues in evaluating employee performance

This study showed that DEA is able to meet the advanced requirement to evaluate employee performance and also addresses some of the major shortcomings of traditional methods. However, the method´s assessment also identified some limitations, including its complexity, the incorporation of so called "soft skills" or its applicability to a small number of employees. In conclusion, the study seems to back up Shields´ (2015) claim that no approach manages to account for all requirements that organizations pose on performance evaluation and the selection of an approach often is a trade-off compromise.

Further, the assessment brought to light, that some general issues on performance evaluation should be considered when assessing employees. This, for instance, includes some conclusions from the Hawthorne experiments. Although many researchers

concluded that there may not be a “Hawthorne effect” in performance evaluations per se (McCambridge et al., 2014), the study showed that there might be specific variables, that are not subject to evaluation, but that affect employees´ performance anyway (e.g. “soft skills”). Also, conclusions from Deming´s Red Beads experiment still apply, considering that individual performance is not only determined by the employee and can be influenced by the system (Burke, 1991). In terms of this research, the sensitivity analysis of DEA´s results showed that the employees´ position in the ranking may vary. In this case, the variation was due to modifications in weights or variables during the sensitivity analysis. This again emphasizes the importance of choosing measures adequately and to consider that they may reflect only parts of performance.

9.4 Limitations and further directions

During data collection, DEA´s application in the case-study context and the subsequent assessment of results, this research encountered some limitations. This chapter will outline the major limitations and also points out future directions of research.

9.4.1 Limitations

The discussion below identifies some of the major limitations under which the study operated.

9.4.1.1 Sample

Although pragmatism and the associated mixed methods approach for data collection and analysis have proven to be an adequate choice to address the research questions, the use of both, qualitative and quantitative methods posed some limitations to this research. First, data collection and analysis required more time than a mono-method approach would probably have. Therefore, the size of the data set for DEA´s application and the sample size of the different groups participating in the assessment could have been larger. Extending the data set, the questionnaire study and the Focus Group interviews would have theoretically been possible. However, the sample size was dependent on the willingness of organizations and individuals to provide data and to voluntarily take part in the assessment and therefore was limited by the scope and resources available to the author.

9.4.1.2 Data collection and analysis

This research also relied on the quality of the empirical data collected, in particular on qualitative data from the Focus Group interviews and quantitative data from the Questionnaire study. Considering the qualitative data, a major limitation lies with the facilitation of the transcription and recording of the interviews. Since the participants did not agree to audio recording, all interviews were documented by taking notes, which were transcribed in comprehensive interview protocol immediately afterwards. However, there is a chance that some of what was being said or other details including voice tones or phrases were lost. The collection of quantitative data, on the other hand, was limited by the kind and number of questions asked. Thus, addressing other topics, including more or modifying some questions could have altered the results. To render transparent questions and results and to offer points of reference to further research, the Focus Groups´ topic guide list, an excerpt of an interview protocol and the full questionnaire are provided in the annex.

Eventually, the analysis of the collected data is limited by the choice of analysis techniques. The Thematic Analysis (TA) that was applied to analyse the Focus Group data provided insightful findings. However, compared to grounded theory or other mature methods, TA lacks sufficient literature which may cause uncertainty in terms of rigorously applying the method. Moreover, TA´s flexibility combining inductive and deductive elements could lead to a lack of consistency when developing themes (Nowell et al., 2017). Considering its flexibility, TA is in coherence with the author´s epistemological position and pragmatic research philosophy. To demonstrate that data analysis has been conducted in a precise, consistent, and exhaustive manner the procedure has been described in detail to enable the reader to determine whether the process is credible. This is also the case for the quantitative analysis which has been carried out by statistical methods including multiple regression and chi-square analysis.

9.4.1.3 Availability of performance data

The application of DEA to evaluate employee performance was also limited considering the availability of data. Thus, the participating bank could only provide data they had previously collected. In consequence, the investigated performance process was restricted to this data, meaning that some performance indicators that would have also been worth investigating (e.g. customer satisfaction per employee) could not be included. Moreover, the data was provided for the group of account managers only so that DEA´s applicability for different job groups could not be assessed and should therefore be investigated in further research.

9.4.1.4 Generalizability of findings

Another major limitation relates to the generalizability of findings from this study. As outlined in chapter 6.2.2, the case study approach often is associated with limited generalization to the population, in particular a single-case approach (Yin, 2013). To strengthen generalizability and to boost confidence in the findings, the organizational context of the case was expanded by involving other groups of managers and workers´ councils than form the particular bank that provided the data for the assessment. Aside from the limited case-study context, the regional context also poses a limitation to this research. Thus, the proposed method was applied in a German organization and assessed by German stakeholders. Previous studies (including Brewster (2007) and Kang and Moon (2011)) showed that there are differences in the way HRM is conducted in different countries. In Germany, as a coordinated market economy (CME), organized interests including workers´ councils play a dominant role and employee rights are rather strong. Hence, organizations in CME´s traditionally have a stronger focus on stakeholder-interests, which may have influenced the choice of performance measures and the subsequent assessment of results. Organizations in liberal market economies like the UK, in contrast, are more sensitive to shareholders´ demands and therefore face high pressure to prioritise short-term shareholder values over employee interests. Also, workers´ councils are rather marginalized (Kang and Moon, 2011). Thus, the design and execution of the performance evaluation in another regional context may have brought up different findings.

Considering that the case study was applied in a narrow context in terms of sample size and regional focus, the author acknowledges that the findings are not generalizable to the population of service employees. Rather, the study may contribute to draw analytic generalizations to inform theory and practices.

9.4.2 Directions for future research

Considering the insights from this research and its limitations, there are some aspects that are considerably worthwhile to investigate in future projects. Thus, this research identified several requirements to performance evaluations to deliver its intended results. Some of them, including “perceived fairness” and the provision of feedback information, could be assessed within this study. However, there are aspects including the induction of (un)desirable behaviour or how employees act on feedback, that only can be assessed after some time. Therefore, the results of the performance evaluation and the decisions

based on it (e.g. promotion, assignment of new territories) should be monitored over a longer period and evaluated eventually.

To address the study´s limited generalizability of findings and availability of data, it seems equitable to expand the study´s context to other service organizations and to other regions. Considering an expansion to other organizations operating in different service contexts (e.g. insurance, tourism, retail) would shed some light on the question whether the proposed method is perceived just as well in other service industries considering its ability to meet general and advanced requirements. Also, it should be examined if its application raises the same issues. Further, an expansion to other service contexts would bring the opportunity to identify possible measures to evaluate the performance of employees with other job profiles and to develop measures to incorporate soft skills. Moreover, the issue of opting for a "subjective component" should be looked at more closely. Thus, it should be investigated, whether a combination of methods is likely to counteract the advantages of DEA or if a combined approach may be superior to the proposed approach.

The study´s expansion to other regions or countries would allow for an analysis on how the economic and cultural context influence the design and perception of HR activities, in particular performance evaluation. Thus, further research should investigate if the whole performance cycle, including the definition of performance measures and the incorporation of different stakeholder interests, differs and subsequently leads to a modified design of the performance evaluation. Moreover, a comparison of the PES´ design in different market economies and its perception by its stakeholders might add some evidence to the current discussion on the critical role of HRM and its contribution to the GFC. Since in some European economies including Germany, the impact of the GFC has not been as intense as in LME´s including the UK and US, there are already some indications that the design of HRM activities may be shaped by the economic context and may cause different perceptions and results. By analysing the similarities and differences regarding the method´s design and stakeholder´s perceptions across different countries, valuable insights on the impact of the economic and cultural context on the critical role of HRM and associated HR activities could be gained.

References

Adler N, Friedman L and Sinuany-Stern Z (2002) Review of ranking methods in the data envelopment analysis context. *European Journal of Operational Research* 140(2): 249-265.

Adler S, Campion M, Colquitt A, Grubb A, Murphy K, Ollander-Krane R, et al. (2016) Getting rid of performance ratings: Genius or folly? A debate. *Industrial and Organizational Psychology* 9(2): 219-252.

Aguinis H (2009) *Performance Management.* : Pearson Prentice Hall Upper Saddle River, NJ.

Aguinis H, Gottfredson RK and Joo H (2012) Delivering effective performance feedback: The strengths-based approach. *Business Horizons* 55(2): 105-111.

Aguinis H, Joo H and Gottfredson RK (2011) Why we hate performance management—And why we should love it. *Business Horizons* 54(6): 503-507.

Albrecht SL, Bakker AB, Gruman JA, Macey WH and Saks AM (2015) Employee engagement, human resource management practices and competitive advantage: An integrated approach. *Journal of Organizational Effectiveness: People and Performance* 2(1): 7-35.

Allen R, Athanassopoulos A, Dyson R and Thanassoulis E (1997) Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions. *Annals of Operations Research* 73: 13-34.

Andersen P and Petersen NC (1993) A procedure for ranking efficient units in data envelopment analysis. *Management Science* 39(10): 1261-1264.

Archer M, Bhaskar R, Collier A, Lawson T and Norrie A (2013) *Critical Realism: Essential Readings.* : Routledge.

Ardichvili A (2013) The role of HRD in CSR, sustainability, and ethics: A relational model. *Human Resource Development Review* 12(4): 456-473.

Ariely D (2013) *What Makes Us Feel Good about our Work? (Video Transcript).*

Armstrong M (2018) *Armstrong's Handbook of Performance Management: An Evidence-Based Guide to Delivering High Performance.* : Kogan Page Publishers.

Armstrong M (2015) *Armstrong's Handbook of Performance Management: An Evidence-Based Guide to Delivering High Performance.* : Kogan Page Publishers.

Armstrong M (2010) *Armstrong's Essential Human Resource Management Practice: A Guide to People Management.* : Kogan Page Publishers.

Armstrong M and Taylor S (2014) *Armstrong's Handbook of Human Resource Management Practice.* : Kogan Page Publishers.

Banker RD and Thrall RM (1992) Estimation of returns to scale using data envelopment analysis. *European Journal of Operational Research* 62(1): 74-84.

Banker RD and Morey RC (1986a) Efficiency analysis for exogenously fixed inputs and outputs. *Operations Research* 34(4): 513-521.

Banker RD and Morey RC (1986b) The use of categorical variables in data envelopment analysis. *Management Science* 32(12): 1613-1627.

Banker RD, Charnes A and Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* 30(9): 1078-1092.

Banker RD, Cooper WW, Seiford LM and Zhu J (2011) Returns to scale in DEA. In: Anonymous *Handbook on Data Envelopment Analysis.* : Springer, 41-70.

Barnabè F (2011) A "system dynamics-based Balanced Scorecard" to support strategic decision making: Insights from a case study. *International Journal of Productivity and Performance Management* 60(5): 446-473.

Barr RS, Seiford LM and Siems TF (1994) Forecasting bank failure: a non-parametric frontier estimation approach. *Recherches Économiques De Louvain/Louvain Economic Review*: 417-429.

Bastiat F (1848/1964) In: Cain S and de Huszar GB (eds) *Selected Essays on Political Economy.* Princeton, NJ: D. van Nordstrand.

Bauer HH, Staat M and Hammerschmidt M (2015) *Grundmodelle Der DEA.* München: Vahlen.

Bayraktar E, Tatoglu E, Turkyilmaz A, Delen D and Zaim S (2012) Measuring the efficiency of customer satisfaction and loyalty for mobile phone brands with DEA. *Expert Systems with Applications* 39(1): 99-106.

Becker B and Gerhart B (1996) The impact of human resource management on organizational performance: Progress and prospects. *Academy of Management Journal* 39(4): 779-801.

Becker BE, Huselid MA, Pickus PS and Spratt MF (1997) HR as a source of shareholder value: Research and recommendations. *Human Resource Management* 36(1): 39-47.

Becker K, Antuar N and Everett C (2011) Implementing an employee performance management system in a nonprofit organization. *Nonprofit Management and Leadership* 21(3): 255-271.

Benkenstein M, Bruhn M, Büttgen M, Hipp C, Matzner M and Nerdinger FW (2017) Topics for Service Management Research-A European Perspective. *SMR-Journal of Service Management Research* 1(1): 4-21.

Berger J, Harbring C and Sliwka D (2013) Performance appraisals and the impact of forced distribution—An experimental investigation. *Management Science* 59(1): 54-68.

Berger AN and Humphrey DB (1997) Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research* 98(2): 175-212.

Birkmeier D, Klöckner S and Overhage S (2015) A survey of service identification approaches-classification framework, state of the art, and comparison. *Enterprise Modelling and Information Systems Architectures* 4(2): 20-36.

Bititci US, Mendibil K, Nudurupati S, Garengo P and Turner T (2006) Dynamics of performance measurement and organisational culture. *International Journal of Operations & Production Management* 26(12): 1325-1350.

Bol JC (2011) The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review* 86(5): 1549-1575.

Boles JS, Donthu N and Lohtia R (1995) Salesperson Evaluation Using Relative Performance Efficiency: The Application of Data Envelopment Analysis. *Journal of Personal Selling & Sales Management* 15(3): 31-49.

Boselie P, Dietz G and Boon C (2005) Commonalities and contradictions in HRM and performance research. *Human Resource Management Journal* 15(3): 67-94.

Boudreau JW and Ramstad PM (1998) Human resource metrics: can measures be strategic? .

Boxall P, Purcell J and Wright P (2007) The goals of HRM. In: Boxall P (ed) *Handbook of Human Resources Management.* : Oxford University Press, 48-67.

Boxall P (2013) Mutuality in the management of human resources: assessing the quality of alignment in employment relationships. *Human Resource Management Journal* 23(1): 3-17.

Bracken DW, Rose DS and Church AH (2016) The evolution and devolution of 360 feedback. *Industrial and Organizational Psychology* 9(4): 761-794.

Bradford DF, Malt RA and Oates WE (1969) The rising cost of local public services: some evidence and reflections. *National Tax Journal* 22(2): 185-202.

Brandon-Jones A, Lewis M, Verma R and Walsman MC (2016) *Examining the Characteristics and Managerial Challenges of Professional Services: An Empirical Study of Management Consultancy in the Travel, Tourism, and Hospitality Sector.*

Braun V and Clarke V (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology* 3(2): 77-101.

Braun V, Clarke V and Terry G (2012) Thematic analysis. *APA Handbook of Research Methods in Psychology* 2: 57-71.

Bremser WG and Wagner WP (2013) Developing dashboards for performance management. *The CPA Journal* 83(7): 62.

Bronzo M, de Resende, Paulo Tarso Vilela, de Oliveira, Marcos Paulo Valadares, McCormack KP, de Sousa PR and Ferreira RL (2013) Improving performance

aligning business analytics with process orientation. *International Journal of Information Management* 33(2): 300-307.

Bryman A (2015) *Social Research Methods.* : Oxford university press.

Budworth M, Latham GP and Manroop L (2015) Looking forward to performance improvement: A field test of the feedforward interview for performance management. *Human Resource Management* 54(1): 45-54.

Burke J (1991) The Parable of the Red Beads. In: Miller RI (ed) *Applying the Deming Method to Higher Education for More Effective Human Resource Management.* : ERIC.

Cappelli P and Tavis A (2016) The Performance Management Revolution. *Harvard Business Review* 94(10): 58-67.

Carifio J and Perla RJ (2007) Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences* 3(3): 106-116.

Cascio WF and Cappelli P (2009) Special Report on Performance Management-Lessons from the Financial Services Crisis-Misguided incentive payouts can push ethical behavior and prudent financial standards out the window. *HR Magazine* 54(1): 46.

Cawley BD, Keeping LM and Levy PE (1998) Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology* 83(4): 615.

Charnes A (1994) *Data Envelopment Analysis: Theory, Methodology and Applications.* : Springer.

Charnes A, Cooper WW, Huang ZM and Sun DB (1990) Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks. *Journal of Econometrics* 46(1): 73-91.

Charnes A, Cooper WW, Golany B, Seiford L and Stutz J (1985) Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics* 30(1): 91-107.

Charnes A, Cooper WW and Rhodes E (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research* 2(6): 429-444.

Chartered Institute of Personnel and Development (2018) *360 Degree Feedback Fact Sheet.*

Chartered Institute of Personnel and Development (2017) *Performance Appraisal Fact Sheet.*

Chartered Institute of Personnel and Development (2016a) Could Do Better: What works in Performance Management . *Research Report*.

Chartered Institute of Personnel and Development (2016b) Rapid evidence assessment of the research literature on the effect of performance appraisal on workplace performance. *Technical Report*.

Chase RB (1978) Where does the customer fit in a service operation? *Harvard Business Review* 56(6): 137-142.

Chytas P, Glykas M and Valiris G (2011) A proactive balanced scorecard. *International Journal of Information Management* 31(5): 460-468.

Clark M (2008) Critical Realim. In: Given LM (ed) *The Sage Encyclopedia of Qualitative Research Methods.* : Sage Publications.

Cocca P and Alberti M (2010) a framework to assess performance measurement systemy in SMEs. *Journal of Productivity and Performance Management* 59(2): 186-200.

Coelli TJ, Rao DSP, O'Donnell CJ and Battese GE (2005) *An Introduction to Efficiency and Productivity Analysis.* : Springer.

Cohen E, Taylor S and Muller-Camen M (2012) HRMs role in corporate social and environmental sustainability. *SHRM Report*.

Cohen J, Cohen P, West SG and Aiken LS (2013) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* : Routledge.

Cook WD and Seiford LM (2009) Data envelopment analysis (DEA) – Thirty years on. *European Journal of Operational Research* 192(1): 1-17.

Cook WD and Zhu J (2006) Rank order data in DEA: A general framework. *European Journal of Operational Research* 174(2): 1021-1038.

Cook WD, Tone K and Zhu J (2014) *Data Envelopment Analysis: Prior to Choosing a Model.*

Cook WD, Kress M and Seiford LM (1996) Data Envelopment Analysis in the Presence of Both Quantitative and Qualitative Factors. *The Journal of the Operational Research Society* 47(7): 945-953.

Cooper WW, Seiford LM and Zhu J (2011) *Handbook on Data Envelopment Analysis.* : Springer.

Cooper WW, Park KS and Yu G (1999) IDEA and AR-IDEA: Models for dealing with imprecise data in DEA. *Management Science* 45(4): 597-607.

Corgnet B, Gómez-Miñambres J and Hernán-Gonzalez R (2015) Goal setting and monetary incentives: When large stakes are not enough. *Management Science* 61(12): 2926-2944.

Coughlan J, Shale E and Dyson R (2010) Including the customer in efficiency analysis: Evidence of a hybrid relational-transactional approach. *International Journal of Bank Marketing* 28(2): 136-149.

Creswell JW and Clark VLP (2007) Designing and conducting mixed methods research.

Dancey CP and Reidy J (2007) *Statistics without Maths for Psychology.* : Pearson Education.

Davenport TH (2006) Competing on analytics. *Harvard Business Review* 84(1): 98.

Delauney J- and Gadrey J (1992) *Services in Economic Thought.* Boston: Kluwer Academic Publishers.

Delery JE and Doty DH (1996) Modes of theorizing in strategic human resource management: Tests of universalistic, contingency, and configurational performance predictions. *Academy of Management Journal* 39(4): 802-835.

Dhungana BR, Nuthall PL and Nartea GV (2004) Measuring the economic inefficiency of Nepalese rice farms using data envelopment analysis. *Australian Journal of Agricultural and Resource Economics* 48(2): 347-369.

Diener E and Crandall R (1978) *Ethics in Social and Behavioral Research.* : U Chicago Press.

Donthu N and Yoo B (1998) Retail Productivity Assessment Using Data Envelopment Analysis. *Journal of Retailing* 74(1): 89-105.

Doyle J and Green RH (1994) Efficiency and Cross-efficiency in DEA: Derivations, Meanings and Uses. *Journal of the Operational Research Society* 45(5): 567.

Du J, Wang J, Chen Y, Chou S and Zhu J (2014) Incorporating health outcomes in Pennsylvania hospital efficiency: an additive super-efficiency DEA approach. *Annals of Operations Research* 221(1): 161-172.

Dulá JH and Hickman BL (1997) Effects of Excluding the Column Being Scored from the DEA Envelopment LP Technology Matrix. *The Journal of the Operational Research Society* 48(10): 1001-1012.

Edvardsson B, Gustafsson A and Roos I (2005) Service portraits in service research: a critical review. *International Journal of Service Industry Management* 16(1): 107-121.

Egbo B (2005) Emergent paradigm: Critical realism and transformative research in educational administration. *McGill Journal of Education* 40(2): 267.

Eken MH, Kale S and Selimler H (2014) Analyzing the Efficiency of European Banks: A DEA-Based Risk and Profitability Approach. In: Anonymous *Global Strategies in Banking and Finance.* : IGI Global, 28-55.

Eker M and Eker S (2009) An empirical analysis of the association between the organizational culture and performance measurement systems in the Turkish manufacturing sector. *Journal of Economic and Social Research* 11(2): 43.

Elliott C and Turnbull S (2004) *Critical Thinking in Human Resource Development.* : Routledge.

Emrouznejad A and Yang G (2017) A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*.

European Banking Federation (2015) *European Banking Sector - Facts and Figures 2015.*

Farquhar J (2004) Customer retention in retail financial services: an employee perspective. *International Journal of Bank Marketing* 22(2): 86.

Farrell MJ (1957) The measurement of productive efficiency. *Journal of the Royal Statistical Society.Series A (General)*: 253-290.

Federal Association of German Banks (2017) *Zahlen, Daten, Fakten Der Kreditwirtschaft.* Berlin: Bundesverband deutscher Banken e.V.

Feilzer Y (2010) Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research* 4(1): 6-16.

Ferrary M (2009) A stakeholder's perspective on human resource management. *Journal of Business Ethics* 87(1): 31-43.

Ferretti F, Saltelli A and Tarantola S (2016) *Trends in Sensitivity Analysis Practice in the Last Decade.*

Fethi MD, Jackson PM and Weyman-Jones TG (2000) Measuring the efficiency of European airlines: an application of DEA and tobit analysis.

Fitzsimmons J and Fitzsimmons M (2008) *Service Management: Operations, Strategy, Information, Technology.* Boston: McGraw Hill.

Fletcher C (2001) Performance appraisal and management: the developing research agenda. *Journal of Occupational and Organizational Psychology* 74(4): 473-487.

Fletcher C (1997) *Appraisal: Routes to Improved Performance.* : Institute of Personnel and Development.

Folan P and Browne J (2005) A review of performance measurement: Towards performance management. *Computers in Industry* 56(7): 663-680.

Foley G (1995) *Understanding Adult Education and Training.* : ERIC.

Freeman RE (1984) Strategic Management: A Stakeholder Approach (Pitman, Boston, MA).

Freeman RE, Harrison JS, Wicks AC, Parmar BL and De Colle S (2010) *Stakeholder Theory: The State of the Art.* : Cambridge University Press.

Freitas H, Oliveira M, Jenkins M and Popjoy O (1998) The focus group, a qualitative research method. *ISRC, Merrick School of Business, University of Baltimore (MD, EUA)* 1: 1-22.

Fried HO, Lovell CK and Schmidt SS (2008) Efficiency and productivity. *The Measurement of Productive Efficiency and Productivity Growth* 3: 3-91.

Gentry JW, Mowen JC and Tasaki L (1991) Salesperson Evaluation: A Systematic Structure for Reducing Judgmental Biases. *Journal of Personal Selling & Sales Management* 11(2): 27.

German Central Bank (1998) *German Banking Act.*

German Federal Bank (2016) *Das Banken- Und Finanzsystem (the Banking and Financial System).*

German Federal Ministry for Economic Affairs and Energy (2015) *Dienstleistungen.*

German Federal Ministry of Justice and Consumer Protection (a) *General Equal Treatment Act.*

German Federal Ministry of Justice and Consumer Protection (b) *Works Consitution Act.*

German Federal Statistical Office (2015) *GDP in Germany 2012 - 2014.*

Glaser BG and Strauss AL (1967) The discovery ofgrounded theory: Strategiesfor qualitative research. *Chicago: Aldire*.

Globerson S (1985) Issues in developing a performance criteria system for an organization. *International Journal of Production Research* 23(4): 639-646.

Godard J (2004) A critical assessment of the high-performance paradigm. *British Journal of Industrial Relations* 42(2): 349-378.

Golany B and Storbeck J (1999) A data envelopment analysis of the operational efficiency of bank branches. *Interfaces* 29(3): 14-26.

Golany B, Learner DB, Phillips FY and Rousseau JJ (1990) Managing service productivity: the data envelopment analysis perspective. *Computers, Environment and Urban Systems* 14(2): 89-102.

Grant D, Keenoy T and Oswick C (1998) Organizational discourse: Of diversity, dichotomy and multi-disciplinarity. *Discourse and Organization*: 1-13.

Greenberg J (1986) Determinants of Percieved Fairness of Performance Evaluations. *Journal of Applied Psychology* 71(2): 340-342.

Grifell Tatjé E and Marques-Gou P (2008) Internal performance evaluation: the case of bank branches. *International Journal of Service Industry Management* 19(3): 302-324.

Grönroos C and Voima P (2013) Critical service logic: making sense of value creation and co-creation. *Journal of the Academy of Marketing Science* 41(2): 133-150.

Grönroos C and Ravald A (2011) Service as business logic: implications for value creation and marketing. *Journal of Service Management* 22(1): 5-22.

Grönroos C and Ojasalo K (2004) Service productivity: Towards a conceptualization of the transformation of inputs into economic results in services. *Journal of Business Research* 57(4): 414-423.

Gruman JA and Saks AM (2011) Performance management and employee engagement. *Human Resource Management Review* 21(2): 123-136.

Guba EG and Lincoln YS (1994a) Competing paradigms in qualitative research. *Handbook of Qualitative Research* 2(163-194): 105.

Guba EG and Lincoln YS (1994b) Competing paradigms in qualitative research. *Handbook of Qualitative Research* 2(163-194): 105.

Guest DE (1997) Human resource management and performance: a review and research agenda. *International Journal of Human Resource Management* 8(3): 263-276.

Guest DE and Conway N (1998) An analysis of the results of the IPD performance management survey. In: Armstrong M and Baron JP (eds) *Performance Management: The New Realities.* : IPD London.

Gummesson E (2014) Productivity, quality and relationship marketing in service operations: A revisit in a new service paradigm. *International Journal of Contemporary Hospitality Management* 26(5): 656-662.

Gupta V and Kumar S (2012) Impact of performance appraisal justice on employee engagement: a study of Indian professionals. *Employee Relations* 35(1): 61-78.

Hailey DJ and Sorgenfrei M (2004) *Measuring Success: Issues in Performance Measurement*.

Hallowell R (1996) The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study. *International Journal of Service Industry Management* 7(4): 27-42.

Harkin B, Webb TL, Chang BP, Prestwich A, Conner M, Kellar I, et al. (2016) Does monitoring goal progress promote goal attainment? A meta-analysis of the experimental evidence. *Psychological Bulletin* 142(2): 198.

Harper S and Vilkinas T (2005) Determining the impact of an organisation's performance management system. *Asia Pacific Journal of Human Resources* 43(1): 76-97.

Harvey D (1990) *The Condition of Postmodernity: An Enquiry into the Conditions of Cultural Change.* : Blackwell.

Hayes RH and Abernathy WJ (2007) Managing our way to economic decline. *Harvard Business Review* 85(7-8).

Haynes P and Fryer G (2000) Human resources, service quality and performance: a case study. *International Journal of Contemporary Hospitality Management* 12(4): 240-248.

Houldsworth E and Jirasinghe D (2006) *Managing and Measuring Employee Performance.* : Kogan Page Publishers.

Howard LW and Miller JL (1993) Fair Pay for Fair Play: Estimating Pay Equity in Professional Baseball with Data Envelopment Analysis. *Academy of Management Journal* 36(4): 882-894.

Hull JC (2012) *Risk Management and Financial Institutions.* : Wiley Finance.

Hutchinson I (2009) *People Glue: Employee Engagement and Retention that Stick.* : Woodslane Press.

Ilgen DR and Feldman JM (1983) Performance appraisal: A process focus. *Research in Organizational Behavior*.

Inman P (2015) *UK Service Sector Rise Boosts Prospect for Economic Growth.*

Jackson Jr. DW, Schlacter JL, Bridges CM and Gallan AS (2010) A Comparison and Expansion of the Bases Used for Evaluating Salespeople's Performance. *Journal of Marketing Theory & Practice* 18(4): 395-406.

Jacobs FR, Chase RB and Aquilano N (2004) Operations management for competitive advantage. *Boston: Mc-Graw Hill* 64: 70.

Jafari M, Bourouni A and Amiri RH (2009) A new framework for selection of the best performance appraisal method. *European Journal of Social Sciences* 7(3): 92-100.

Jamieson S (2004) Likert scales: how to (ab) use them. *Medical Education* 38(12): 1217-1218.

Jawahar I (2010) The mediating role of appraisal feedback reactions on the relationship between rater feedback-related behaviors and ratee performance. *Group & Organization Management* 35(4): 494-526.

Jeffrey SA, Schulz A and Webb A (2012) The performance effects of an ability-based approach to goal assignment. *Journal of Organizational Behavior Management* 32(3): 221-241.

Johnson RB and Onwuegbuzie AJ (2004) Mixed methods research: A research paradigm whose time has come. *Educational Researcher* 33(7): 14-26.

Kahn WA (1990) Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal* 33(4): 692-724.

Kahneman D and Tversky A (1977) *Intuitive Prediction: Biases and Corrective Procedures*.

Kampkötter P and Sliwka D (2011) Differentiation and performance: An empirical investigation on the incentive effects of bonus plans.

Kanji GK and e Sá PM (2002) Kanji's business scorecard. *Total Quality Management* 13(1): 13-27.

Kaplan RS and Norton DP (1992) The Balanced Scorecard--Measures That Drive Performance. *Harvard Business Review* 70(1): 71-79.

Keeble-Ramsay DR and Armitage A (2015) HRD challenges faced in the post-global financial crisis period–insights from the UK. *European Journal of Training and Development* 39(2): 86-103.

Keeble-Ramsay DR and Armitage A (2010) Total quality management meets human resource management: Perceptions of the shift towards high performance working. *The TQM Journal* 22(1): 5-25.

Kellough JE and Nigro LG (2002) Pay for performance in Georgia state government: Employee perspectives on GeorgiaGain after 5 years. *Review of Public Personnel Administration* 22(2): 146-166.

Kemble R and Keeble-Ramsay D (2014) Conceptual analysis of post global credit crunch and impact for employee engagement within the UK.

Kim T and Holzer M (2016) Public employees and performance appraisal: A study of antecedents to employees' perception of the process. *Review of Public Personnel Administration* 36(1): 31-56.

Kitzinger J (1994) The methodology of focus groups: the importance of interaction between research participants. *Sociology of Health & Illness* 16(1): 103-121.

Kitzinger J (1995) Qualitative research. Introducing focus groups. *BMJ (Clinical Research Ed.)* 311(7000): 299-302.

Kline TJB and Sulsky LM (2009) Measurement and assessment issues in performance appraisal. *Canadian Psychology/Psychologie Canadienne* 50(3): 161-171.

Kluger AN and DeNisi A (1996) The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 119(2): 254.

Knox Lovell CA and Pastor JT (1997) Target setting: An application to a bank branch network. *European Journal of Operational Research* 98(2): 290-299.

Kochanski J (2007) Sibson reveals secrets of successful performance management. *Employee Benefit News*: 22-23.

Koch-Rogge M, Westermann G, Wilbert C and Willis R (2014) How are they doing? The Application of Data Envelopment Analysis to Develop a Strategic Performance Appraisal on Individual Level. *A Focused Issue on Building New Competences in Dynamic Environments (Research in Competence-Based Management, Volume 7) Emerald Group Publishing Limited* 7: 193-218.

Kondrasuk J (2012) The ideal performance appraisal is a format, not a form. *Academy of Strategic Management Journal* 11(1): 115.

Koopmans TC (1951) Analysis of production as an efficient combination of activities. *Activity Analysis of Production and Allocation* 13: 33-37.

Kramar R (2014) Beyond strategic human resource management: is sustainable human resource management the next approach? *The International Journal of Human Resource Management* 25(8): 1069-1089.

Krueger R and Casey MA (2015) *Focus Groups: A Practical Guide for Applied Research.* : Sage publications.

Kuckartz U, Rädiker S, Ebert T and Schehl J (2010) Kreuztabelle, Chi-Quadrat und Zusammenhangsmaße. In: Anonymous *Statistik.* : Springer, 81-102.

Kuhn TS (1969) The Structure of Scientific Revolutions, Postscript. *Book*.

LaPlante AE and Paradi J (2015) Evaluation of bank branch growth potential using data envelopment analysis. *Omega* 52: 33-41.

Latham GP (2004) The motivational benefits of goal-setting. *The Academy of Management Executive* 18(4): 126-129.

Latham GP and Locke EA (2006) Enhancing the benefits and overcoming the pitfalls of goal setting. *Organizational Dynamics* 35(4): 332-340.

Latham GP, Skarlicki D, Irvine D and Siegel JP (1993) The increasing importance of performance appraisals to employee effectiveness in organizational settings in North America. *International Review of Industrial and Organizational Psychology* 8: 87-132.

Lee BL and Worthington AC (2016) A network DEA quantity and quality-orientated production model: An application to Australian university research services. *Omega* 60: 26-33.

Lee J and Lee H (2014) *Developing and Validating a Citizen-Centric Typology for Smart City Services.*

Leibenstein H and Maital S (1992) Empirical estimation and partitioning of X-inefficiency: a data-envelopment approach. *The American Economic Review* 82(2): 428-433.

Leshem S and Trafford V (2007) Overlooking the conceptual framework. *Innovations in Education and Teaching International* 44(1): 93-105.

Levy PE and Williams JR (2004) The social context of performance appraisal: A review and framework for the future. *Journal of Management* 30(6): 881-905.

Lipset , S. M. , Trow , M. , & Coleman , J. (1956) *Union Democracy: The Inside Politics of the International Typographical Union.* New York.

Liu JS, Lu LYY and Lu W (2016) *Research Fronts in Data Envelopment Analysis.*

Lovell CK and Pastor JT (1995) Units invariant and translation invariant DEA models. *Operations Research Letters* 18(3): 147-151.

Lovelock C and Gummesson E (2004) Whither services marketing? In search of a new paradigm and fresh perspectives. *Journal of Service Research* 7(1): 20-41.

Lovelock CH (1991) Distinctive aspects of services marketing. *Lovelock CH.Services Marketing.Philadelphia.Prentice-Hall International Inc*: 1-11.

Lubke GH and Muthén BO (2004) Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling* 11(4): 514-534.

Lusch R (1990) Personal differences, job tension, job outcomes, and store performance : a study of retail store managers. *Journal of Marketing* 54(1).

Lynch RL and Cross KF (1991) *Measure Up!: The Essential Guide to Measuring Business Performance.* : Mandarin.

MacKenzie CA, Garavan TN and Carbery R (2012) Through the looking glass: challenges for human resource development (HRD) post the global financial crisis – business as usual? *Human Resource Development International* 15(3): 353-364.

Maister DH and Lovelock CH (1982) Managing facilitator services. *Sloan Management Review* 23(4): 19.

Malos SB (1998) Current legal issues in performance appraisal. *Performance Appraisal: State of the Art in Practice*: 49-94.

Mani BG (2002) *Performance Appraisal Systems, Productivity, and Motivation: A Case Study.*

Manoharan TR, Muralidharan C and Deshmukh SG (2009) Employee Performance Appraisal Using Data Envelopment Analysis: A case study. *Research and Practice in Human Resource Management* 17(1): 92-111.

Mariappanadar S (2014) *Stakeholder Harm Index: A Framework to Review Work Intensification from the Critical HRM Perspective.*

Martinez V, Radnor Z, Radnor ZJ and Barnes D (2007) Historical analysis of performance measurement and management in operations management. *International Journal of Productivity and Performance Management* 56(5/6): 384-396.

Maurer I and Müller-Camen M (2016) Nachhaltiges Personalmanagement. In: Anonymous *CSR Und Human Resource Management.* : Springer, 17-30.

Mayo E (1933) The human problems of an industrial organization. *New York: McMillan*.

McCambridge J, Witton J and Elbourne DR (2014) Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of Clinical Epidemiology* 67(3): 267-277.

Menefee JA and Murphy RO (2004) Rewarding and retaining the best: Compensation strategies for top performers. *Benefits Quarterly* 20(3): 13.

Mertens DM (2014) *Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed Methods.* : Sage publications.

Messersmith JG, Patel PC, Lepak DP and Gould-Williams JS (2011) Unlocking the black box: exploring the link between high-performance work systems and performance. *Journal of Applied Psychology* 96(6): 1105.

Meyer HH (1991) A solution to the performance appraisal feedback enigma. *Executive (19389779)* 5(1): 68-76.

Micheli P and Kennerley M (2005) Performance measurement frameworks in public and non-profit sectors. *Production Planning & Control* 16(2): 125-134.

Micheli P and Mari L (2014) The theory and practice of performance measurement. *Management Accounting Research* 25(2): 147-156.

Moreno AA and Tadepalli R (2002) Assessing academic department efficiency at a public university. *Managerial and Decision Economics* 23(7): 385-397.

Morgan DL (2007) Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research* 1(1): 48-76.

Murphy KR and Cleveland J (1995) *Understanding Performance Appraisal: Social, Organizational, and Goal-Based Perspectives.* : Sage.

Myers JL, Well A and Lorch RF (2010) *Research Design and Statistical Analysis.* : Routledge.

Nazarko J and Šaparauskas J (2014) Application of DEA method in efficiency evaluation of public higher education institutions. *Technological and Economic Development of Economy* 20(1): 25-44.

Neely A (2005) The evolution of performance measurement research: developments in the last decade and a research agenda for the next. *International Journal of Operations & Production Management* 25(12): 1264-1277.

Neely A and Austin R (2000) Measuring operations performance: past, present and future. *Performance Measurement*: 419-426.

Neely A, Adams C and Crowe P (2001) The performance prism in practice. *Measuring Business Excellence* 5(2): 6-13.

Neely A, Gregory M and Platts K (1995) Performance measurement system design: a literature review and research agenda. *International Journal of Operations & Production Management* 15(4): 80-116.

Neely A, Richards H, Mills J, Platts K and Bourne M (1997) Designing performance measures: a structured approach. *International Journal of Operations & Production Management* 17(11): 1131-1152.

Neely A, Mills J, Platts K, Richards H, Gregory M, Bourne M, et al. (2000) Performance measurement system design: developing and testing a process-based approach. *International Journal of Operations & Production Management* 20(10): 1119-1145.

Neubert MJ (1998) The value of feedback and goal setting over goal setting alone and potential moderators of this effect: A meta-analysis. *Human Performance* 11(4): 321-335.

Noor KBM (2008) Case study: A strategic research methodology. *American Journal of Applied Sciences* 5(11): 1602-1604.

Nowell LS, Norris JM, White DE and Moules NJ (2017) Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods* 16(1): 1609406917733847.

Nudurupati SS, Bititci US, Kumar V and Chan FTS (2011) *State of the Art Literature Review on Performance Measurement.*

O'Donnell D, McGuire D and Cross C (2006) Critically challenging some assumptions in HRD. *International Journal of Training and Development* 10(1): 4-16.

Ostrom AL, Parasuraman A, Bowen DE, Patricio L and Voss CA (2015) Service research priorities in a rapidly changing context. *Journal of Service Research* 18(2): 127-159.

Paauwe J, Wright P and Guest D (2013) HRM and performance: What do we know and where should we go. *HRM and Performance: Achievements and Challenges*: 1-13.

Paradi JC, Vela SA and Zhu H (2010) Adjusting for cultural differences, a new DEA model applied to a merged bank. *Journal of Productivity Analysis* 33(2): 109-123.

Paradi JC, Smith S and Schaffnit C (2006) Knowledge worker performance analysis using DEA: an application to engineering design teams at Bell Canada. *IEEE Transactions on Engineering Management* 49(2): 161-172.

Paradi JC and Zhu H (2013) A survey on bank branch efficiency and performance research with data envelopment analysis. *Omega* 41(1): 61-79.

Paradi JC and Schaffnit C (2004) Commercial branch performance evaluation and results communication in a Canadian bank—a DEA application. *European Journal of Operational Research* 156(3): 719-735.

Paradi JC, Rouatt S and Zhu H (2011) Two-stage evaluation of bank branch efficiency using data envelopment analysis. *Omega* 39(1): 99-109.

Parasuraman A, Zeithaml VA and Berry LL (1988) Servqual. *Journal of Retailing* 64(1): 12-40.

Pedraja-Chaparro F, Salinas-Jimenez J and Smith P (1997) On the role of weight restrictions in data envelopment analysis. *Journal of Productivity Analysis* 8(2): 215-230.

Penrose ET (1959) *The Growth of the Firm.* White Plains, NY: Sharpe.

Perry C, Riege A and Brown L (1999) Realism's role among scientific paradigms in marketing research. *Irish Marketing Review* 12(2): 16.

Pettijohn LS, Parker RS and Pettijohn CE (2001) Performance appraisals: usage, criteria and observations. *Journal of Management Development* 20(9): 754-771.

Pettijohn CE, Pettijohn LS and d'Amico M (2001) Characteristics of Performance Appraisals and Their Impact on Sales Force Satisfaction. *Human Resource Development Quarterly* 12(2): 127-146.

Pilling BK, Donthu N and Henson S (1999) Accounting for the Impact of Territory Characteristics on Sales Performance: Relative Efficiency as a Measure of Salesperson Performance. *Journal of Personal Selling & Sales Management* 19(2): 35-45.

Pleier N (2008) Performance-measurement-systeme und der faktor mensch. *Aufl., Wiesbaden*.

Pohl J (2015) Performance Management: Dogma oder Einzelfallbetrachtung? In: Künzel H (ed) *Erfolgsfaktor Performance Management: Leistungsbereitschaft Einer Aufgeklärten Generation.* : Springer-Verlag, 1-16.

Post JE, Preston LE and Sauter-Sachs S (2002) *Redefining the Corporation: Stakeholder Management and Organizational Wealth.* : Stanford University Press.

Pricewaterhouse Coopers (2011) *Effektives Kundenmanagement Im Retail Banking.* München: Pricewaterhouse Coopers AG.

Psychogois A, Brewster C, Parry E, Dickmann M, Brewster C and Sparrow P (2016) 7 Western European HRM: Reactions and Adjustment to Crises. *International Human Resource Management: Contemporary HR Issues in Europe*: 115.

Pun KF and White AS (2005) A performance measurement paradigm for integrating strategy formulation: A review of systems and frameworks. *International Journal of Management Reviews* 7(1): 49-71.

Purcell J, Hutchinson S, Kinnie N, Rayton B and Swart J (2003) Understanding the pay and performance link: Unlocking the black box. *London: CIPD*.

Rabiee F (2004) Focus-group interview and data analysis. *Proceedings of the Nutrition Society*: 655-660.

Rathmell JM (1966) What is meant by services? *Journal of Marketing*(30): 32-60.

Roberts GE (2003) Employee performance appraisal system participation: A technique that works. *Public Personnel Management* 32(1): 89-98.

Ruggiero J (1998) Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research* 111(3): 461-469.

Rust RT and Huang M (2012) Optimizing service productivity. *Journal of Marketing* 76(2): 47-66.

Sambrook S (2004) A “critical” time for HRD? *Journal of European Industrial Training* 28(8/9): 611-624.

Saunders M, Lewis P and Thornhill A (2012) *Research Methods for Business Students.* : Pearson.

Savaneviciene A and Stankeviciute Z (2010) The Models Exploring the “Black Box” between HRM and Organizational Performance. *Inzinerine Ekonomika-Engineering Economics* 21(4): 426-434.

Schaffnit C, Rosen D and Paradi JC (1997) Best practice analysis of bank branches: An application of DEA in a large Canadian bank. *European Journal of Operational Research* 98(2): 269-289.

Schläfke M, Silvi R and Möller K (2012) A framework for business analytics in performance management. *International Journal of Productivity and Performance Management* 62(1): 110-122.

Schmenner RW (2004) Service businesses and productivity. *Decision Sciences* 35(3): 333-347.

Schmenner RW (1986) How can service businesses survive and prosper? *Sloan Management Review (1986-1998)* 27(3): 21.

Schneider B (1994) HRM-a service perspective: towards a customer-focused HRM. *International Journal of Service Industry Management* 5(1): 64-76.

Sebald H and Jacob L (2015) Die Etablierung einer Leistungskultur durch die konsequente Nutzung von Performance Management. In: Anonymous *Change Management in Versicherungsunternehmen.* : Springer, 193-213.

Seiford LM and Zhu J (1999a) Infeasibility of super-efficiency data envelopment analysis models. *Infor* 37(2): 174-187.

Seiford LM and Thrall RM (1990) Recent developments in DEA: the mathematical programming approach to frontier analysis. *Journal of Econometrics* 46(1): 7-38.

Seiford LM and Zhu J (1999b) Profitability and Marketability of the Top 55 U.S. Commercial Banks. *Management Science* 45(9): 1270-1288.

Sexton TR, Silkman RH and Hogan AJ (1986) Data envelopment analysis: Critique and extensions. *New Directions for Program Evaluation* 1986(32): 73-105.

Sherman HD and Ladino G (1995) Managing bank productivity using data envelopment analysis (DEA). *Interfaces* 25(2): 60-73.

Shields J, Brown M, Kaine S, Dolle-Samuel C, North-Samardzic A, McLean P, et al. (2015) *Managing Employee Performance & Reward: Concepts, Practices, Strategies.* : Cambridge University Press.

Shirouyehzad H, Hosseinzadeh Lofti F, Aryanezhad MB and Dabestani R (2012) A DATA ENVELOPMENT ANALYSIS APPROACH FOR MEASURING THE EFFICIENCY OF EMPLOYEES: A CASE STUDY. *South African Journal of Industrial Engineering* 23(1).

Silverman M, Kerrin M and Carter A (2005) *360 Degree Feedback: Beyond the Spin.* : Institute for Employment Studies.

Silvestro R (1999) Positioning services along the volume-variety diagonal: the contingencies of service design, control and improvement. *International Journal of Operations & Production Management* 19(4): 399-421.

Silvestro R, Fitzgerald L, Johnston R and Voss C (1992) Towards a classification of service processes. *International Journal of Service Industry Management* 3(3): 62-75.

Smith S (2013) Employee Performance Appraisal Process: Honesty is the Best Policy.

Smither JW, London M and Reilly RR (2005) Does performance improve following multisource feedback? A theoretical model, meta-analysis, and review of empirical findings. *Personnel Psychology* 58(1): 33-66.

Smither J (1998) Lessons Learned: Research Implications for Performance Appraisal and Management. In: Smither J (ed) *Performance Appraisals: A State of the Art in Practice.* San Francisco: Jossey-Bass San Francisco, CA, 537-547.

Soteriou A and Zenios SA (1999) Operations, quality, and profitability in the provision of banking services. *Management Science* 45(9): 1221-1238.

Soteriou AC and Stavrinides Y (1997) *An Internal Customer Service Quality Data Development Analysis Model for Bank Branches.*

Star S, Russ-Eft D, Braverman MT and Levine R (2016) Performance Measurement and Performance Indicators: A Literature Review and a Proposed Model for Practical Adoption. *Human Resource Development Review* 15(2): 151-181.

Storey J (2007) *Human Resource Management: A Critical Text.* : Cengage Learning EMEA.

Sudin S (2011) Fairness of and satisfaction with performance appraisal process. *Journal of Global Management* 2(1): 66-83.

Sue VM and Ritter LA (2011) *Conducting Online Surveys.* : SAGE publications.

Sureshchandar G and Leisten R (2005) Holistic scorecard: strategic performance measurement and management in the software industry. *Measuring Business Excellence* 9(2): 12-29.

Takeuchi R, Chen G and Lepak DP (2009) Through the looking glass of a Social System: Cross-level effects of High-Performance Work Systems on Employees´ Attitudes. *Personnel Psychology* 62(1): 1-29.

Tan Y, Zhang Y and Khodaverdi R (2017) Service performance evaluation using data envelopment analysis and balance scorecard approach: an application to automotive industry. *Annals of Operations Research* 248(1-2): 449-470.

Tang TL and Sarsfield-Baldwin LJ (1996) Distributive and procedural justice as related to satisfaction and commitment. *SAM Advanced Management Journal* 61: 25-31.

Tangen S (2004) *Evaluation and Revision of Performance Measurement Systems.* Stockholm: Department of Production Engineering, The Royal Institute of Technology.

Taticchi P, Tonelli F and Cagnazzo L (2010) Performance measurement and management: a literature review and a research agenda. *Measuring Business Excellence* 14(1): 4-18.

Tavakol M and Dennick R (2011) Making sense of Cronbach's alpha. *International Journal of Medical Education* 2: 53.

Thanassoulis E, Portela MC and Allen R (2004) Incorporating value judgments in DEA. In: Anonymous *Handbook on Data Envelopment Analysis.* : Springer, 99-138.

Thanassoulis E (1993) A Comparison of Regression Analysis and Data Envelopment Analysis as Alternative Methods for Performance Assessments. *The Journal of the Operational Research Society* 44(11): 1129-1144.

Thanassoulis E and Dyson RG (1992) Estimating preferred target input-output levels using data envelopment analysis. *European Journal of Operational Research* 56(1): 80-97.

Thanassoulis E (1995) Assessing police forces in England and Wales using data envelopment analysis. *European Journal of Operational Research* 87(3): 641-657.

The Human Resources Social Network (2001) *Managing Processes Instead of People.*

Thomas BE and Baron JP (1994) *Evaluating Knowledge Worker Productivity: Literature Review*.

Thomas SL and Bretz RD (1994) Research and practice in performance appraisal: evaluating employee performance in America´s largest companies. *SAM Advanced Management Journal* 1994 Spring(22 (2)): 28-37.

Thomas RR, Barr RS, Cron WL and Slocum Jr. JW (1998) A process for evaluating retail store efficiency: a restricted DEA approach. *International Journal of Research in Marketing* 15(5): 487-503.

Thompson P (2011) The trouble with HRM. *Human Resource Management Journal* 21(4): 355-367.

Thompson RG, Singleton Jr F, Thrall RM and Smith BA (1986) Comparative site evaluations for locating a high-energy physics lab in Texas. *Interfaces* 16(6): 35-49.

Thor C (1993) A complete productivity and quality measurement system. *Handbook for Productivity Measurement and Improvement, Productivity Press, Cambridge, MA*.

Thoraneenitiyan N and Avkiran NK (2009) Measuring the impact of restructuring and country-specific factors on the efficiency of post-crisis East Asian banking systems: Integrating DEA with SFA. *Socio-Economic Planning Sciences* 43(4): 240-252.

Thrall RM (1996) Duality, classification and slacks in DEA. *Annals of Operations Research* 66(2): 109-138.

Timmer CP (1971) Using a probabilistic frontier production function to measure technical efficiency. *The Journal of Political Economy*: 776-794.

Tinnilä M (2012) A Review of service frameworks analyzing strategic repositioning: the case of bank services. *Implementation and Integration of Information Systems in the Service Sector*: 1.

Tone K (2001) A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research* 130(3): 498-509.

Tzeremes N and Halkos G (2010) A DEA approach for measuring university departments' efficiency.

Van Maanen J, Sørensen JB and Mitchell TR (2007) The interplay between theory and method. *Academy of Management Review* 32(4): 1145-1154.

Vargo SL and Lusch RF (2008) Why "service"? *Journal of the Academy of Marketing Science* 36(1): 25-38.

Vargo SL and Lusch RF (2004) The four service marketing myths remnants of a goods-based, manufacturing model. *Journal of Service Research* 6(4): 324-335.

Varma A, Denisi AS and Peters LH (1996) Interpersonal affect and performance appraisal: A field study. *Personnel Psychology* 49(2): 341-360.

Vater D, Cho Y and Sidebottom P (2012) *Retail-Banking: Die Digitale Herausforderung.*

Verweir K (2006) Integrated performance management: A guide to strategy implementation. *Strategic Direction* 22(8).

Wagner JM, Shimshak DG and Novak MA (2003) Advances in physician profiling: the use of DEA. *Socio-Economic Planning Sciences* 37(2): 141-163.

Wall TD and Wood SJ (2005) The romance of human resource management and business performance, and the case for big science. *Human Relations* 58(4): 429-462.

Walras L (1894/ 1954) *Elements of the Political Economy (Reprint).* Homestead, NJ: Richard D. Irwin.

Wang Y and Chin K (2010) Some alternative models for DEA cross-efficiency evaluation. *International Journal of Production Economics* 128(1): 332-338.

Wanke P and Barros C (2016) Efficiency in Latin American airlines: a two-stage approach combining Virtual Frontier Dynamic DEA and Simplex Regression. *Journal of Air Transport Management* 54: 93-103.

Warning S (2014) How to pick your staff? Using data envelopment analysis. *Management Research Review* 37(9): 815-832.

Weathington BL, Cunningham CJ and Pittenger DJ (2010) *Research Methods for the Behavioral and Social Sciences.* : John Wiley & Sons.

Wells S (2011) HRM for sustainability: Creating a new paradigm. In: Clarke M (ed) *Readings in HRM and Sustainability.* : Tilde University Press.

Welsh DT and Ordóñez LD (2014) Conscience without cognition: The effects of subconscious priming on ethical behavior. *Academy of Management Journal* 57(3): 723-742.

Westermann G (1996) *Lokaler Technischer Fortschritt Und Intra-Industrielle Strukturen: Eine Data-Envelopment-Analyse.* : Shaker.

Westermann G and Johnson G (1999) Combining DEA and “Transformation-Stages”: Management Strategies for the Disability Service Units of the St. Georg Association. *Data Envelopment Analysis in the Service Sector, Wiesbaden*.

Wirtz J and Ehret M (2017) Capturing Value in the Service Economy. *SMR-Journal of Service Management Research* 1(1): 22-38.

Wright PM and Nishii LH (2007) Strategic HRM and organizational behavior: Integrating multiple levels of analysis. *CAHRS Working Paper Series*: 468.

Yadav N and Sagar M (2013) Performance measurement and management frameworks: Research trends of the last two decades. *Business Process Management Journal* 19(6): 947-971.

Yeates K (2015) *Motivating and Retaining Employees.*

Yen I, Bastani F, Huang Y, Zhang Y and Yao X (2017) *SaaS for Automated Job Performance Appraisals using Service Technologies and Big Data Analytics.* : IEEE.

Yin RK (2017) *Case Study Research and Applications: Design and Methods.* : Sage publications.

Yin RK (2013) Validity and generalization in future case study evaluations. *Evaluation* 19(3): 321-332.

Zeithaml V and Bitner M (2003) Services Marketing 3rd edition New York The McGraw Hill Company.

Zeithaml VA, Parasuraman A and Berry LL (1985) Problems and strategies in services marketing. *The Journal of Marketing*: 33-46.

Appendices

Annex 1: Data summary sheets

Data Summary

stage: acquisition

DEA model = CCR-O

No. of DMUs = 40
No. Input items = 3
Input(1) = CUST
Input(2) = INHAB
Input(3) = VOLACT
No. of Output items = 2
Output(1) = SELL
Output(2) = APP

Returns to Scale = Constant (0 =< Sum of Lambda < Infinity)

Statistics on Input/Output Data

	CUST	INHAB	VOLACT	SELL	APP
Min	111.90	2154.00	72395.50	1.00	65.00
Max	1648.50	39784.00	22319537.01	807.00	716.00
Mean	680.77	19081.13	2679019.41	160.00	342.98
SD	360.32	9936.90	5228208.20	116.89	149.98

Correlation

	CUST	INHAB	VOLACT	SELL	APP
CUST	1.000	-0.341	-0.496	-0.097	0.379
INHAB		1.000	0.273	-0.033	-0.120
CUSVOL			1.000	0.055	-0.061
SELLQ				1.000	0.305
APPQ					1.000

DMUs with inappropriate Data with respect to the chosen Model

No.	DMU
	None

Data Summary

stage: consultancy and sales

DEA model = CCR-O

No. of DMUs = 40
No. Input items = 2
Input(1) = SELL
Input(2) = APP
No. of Output items = 5
Output(1) = VOLACT
Output(2) = VOLL
Output(3) = BSH
Output(4) = RuV
Output(5) = Union

Returns to Scale = Constant (0 =< Sum of Lambda < Infinity)

Statistics on Input/Output Data

	SELL	APP	VOLACT	VOLL	BSH	RuV	Union
Min	1.00	65.00	72395.50	169933.92	21616.21	29825.00	3416.00
Max	807.00	716.00	22319537.01	21680434.59	1483526.77	1057728.00	3715129.00
Mean	160.00	342.98	2679019.41	5576449.59	515614.02	303687.08	682826.16
SD	116.89	149.98	5228208.20	4139354.65	303895.86	301696.83	826452.73

Correlation

	SELL	APP	VOLACT	VOLL	BSH	RuV	Union
SELLQ	1.000	0.305	0.055	0.364	0.207	0.487	0.352
APPQ		1.000	-0.061	0.175	0.567	0.203	0.488
VOLACT			1.000	0.150	-0.267	0.195	-0.184
VOLL				1.000	0.499	**0.763**	0.604
BSH					1.000	0.410	0.533
RuV						1.000	0.598
Union							1.000

DMUs with inappropriate Data with respect to the chosen Model

No.	DMU
	None

Data Summary

stage: profitability

DEA model = CCR-O

No. of DMUs = 40

No. Input items = 3

Input(1) = SAL

Input(2) = TRAI

Input(3) = BONUS

No. of Output items = 2

Output(1) = CMII

Output(2) = CommE

Returns to Scale = Constant (0 =< Sum of Lambda < Infinity)

Statistics on Input/Output Data

	SAL	Bonus	TRAI	CMII	CommE
Min	18259.33	1.00	1.00	17381.71	5474.20
Max	86834.14	14000.00	17126.95	406278.75	102435.89
Mean	45252.66	1525.43	2190.42	114967.76	31758.12
SD	14319.87	3028.71	3690.01	94762.98	26054.88

Correlation

	SAL	Bonus	TRAI	CMII	CommE
SAL	1.000	0.738	0.743	0.670	0.721
Bonus		1.000	0.986	0.397	0.742
TRAI			1.000	0.397	0.553
CMII				1.000	0.742
CommE					1.000

DMUs with inappropriate Data with respect to the chosen Model

No.	DMU
	None

Data Summary Sheet

stage: bonuses

DEA model = CCR-O/ CCR-I

No. of DMUs = 40

No. Input items = 2 (3 for CCR-I)

Input(1) = SAL

Input(2) = BONUS

Input(3) = BONUS EXP. (CCR-I only)

No. of Output items = 2

Output(1) = CMII

Output(2) = CommE

Returns to Scale = Constant (0 =< Sum of Lambda < Infinity)

Statistics on Input/Output Data

	SAL	CMII	CommE	Bonus	Bonus exp. (CCR-I only)
Min	18259.33	17381.71	5474.20	1.00	1925.00
Max	86834.14	406278.75	102435.89	14000.00	1925.00
Mean	45252.66	114967.76	31758.12	1525.43	1925.00
SD	14319.87	94762.98	26054.88	3028.71	0.00

Correlation

	SAL	CMII	CommE	Bonus
SAL	1.00	0.67	0.72	0.74
CMII		1.00	0.74	0.41
CommE			1.00	0.54
Bonus				1.00

DMUs with inappropriate Data with respect to the chosen Model

No.	DMU
	None

Data Summary

stage: performance targets

DEA model = CCR-O

No. of DMUs = 40
No. Input items = 5
Input(1) = CUST
Input(2) = INHAB
Input(3) = VOLACT
Input(4) = SAL
No. of Output items = 2
Output(1) = CMII
Output(2) = CommE

Returns to Scale = Constant (0 =< Sum of Lambda < Infinity)

Statistics on Input/Output Data

	CUST	INHAB	VOLACT	SAL	CMII	CommE
Min	111.90	2154.00	72395.50	18259.33	17381.71	5474.20
Max	1648.50	39784.00	22319537.01	86834.14	406278.75	102435.89
Mean	680.77	19081.13	2679019.41	45252.66	114967.76	31758.12
SD	360.32	9936.90	5228208.20	14319.87	94762.98	26054.88

Correlation

	CUST	INHAB	VOLACT	SAL	CMII	CommE
CUST	1.000	-0.280	-0.496	-0.290	-0.387	-0.474
INHAB		1.000	0.272	0.196	0.277	0.327
VOLACT			1.000	0.554	0.778	0.628
SAL				1.000	0.670	0.721
CMII					1.000	0.742
CommE						1.000

DMUs with inappropriate Data with respect to the chosen Model

No.	DMU
	None

Sensitivity Analysis Category	1	weights imposed
	2	variable(s) removed
	3	Superefficiency
	4	DMU(s) removed

TS 1 “acquisition”

	No restrictions	weights imposed (all inputs and outputs min 10%)	superefficiency	DMU removed (DMU 34)
Sensitivity Analysis Category		1	3	4
Unit name	Score	Score	Score	Score
1	0.68	0.55	0.68	0.68
2	1.00	1.00	1.05	1.00
3	0.66	0.66	0.66	0.66
4	0.67	0.62	0.67	0.68
5	0.67	0.62	0.67	0.68
6	0.92	0.83	0.92	0.92
7	0.37	0.31	0.37	0.37
8	0.53	0.49	0.53	0.53
9	1.00	1.00	1.17	1.00
10	0.91	0.85	0.91	0.91
11	0.64	0.59	0.64	0.64
12	1.00	0.83	1.26	1.00
13	0.99	0.78	0.99	0.99
14	0.80	0.71	0.80	0.80
15	0.46	0.46	0.46	0.46
16	0.60	0.54	0.60	0.64
17	0.78	0.69	0.78	0.78
18	0.35	0.33	0.35	0.36
19	0.78	0.71	0.78	0.79
20	1.00	1.00	1.43	1.00
21	0.86	0.76	0.86	0.86
22	0.52	0.48	0.52	0.54
23	0.48	0.43	0.48	0.48
24	0.82	0.78	0.82	0.85
25	1.00	1.00	1.09	1.00
26	0.65	0.49	0.65	1.00
27	0.41	0.40	0.41	0.41
28	0.93	0.76	0.93	1.00
29	0.40	0.01	0.40	0.40
30	0.90	0.83	0.90	0.90
31	0.69	0.69	0.69	0.69
32	0.91	0.80	0.91	0.99
33	0.31	0.26	0.31	0.31
34	1.00	1.00	3.11	

35	0.38	0.37	0.38	0.66
36	1.00	1.00	1.32	1.00
37	1.00	1.00	1.06	1.00
38	1.00	1.00	1.62	1.00
39	1.00	1.00	1.19	1.00
40	1.00	0.99	1.08	1.00

MIN	0.31	0.01	0.31	0.31
MAX	1.00	1.00	3.11	1.00
MEAN	0.75	0.69	0.86	0.78
SD	0.23	0.25	0.48	0.23
Number efficient units	11.00	9.00	0.00	12.00

TS 2 “consultancy and sales”

	no restrictions	weights imposed	variable removed (RuV)	super-efficiency	DMU removed (no. 26)	DMU removed (no. 29)	DMU removed (no. 7)	DMUs (no.7, 26 and 29) and variable (RuV) removed
Sensitivity Analysis Category		1	2	3	4	4	4	2,4
Unit name	Score	Score	Score	Score	Score	Score	Score	Score
1	0.35	0.18	0.35	0.35	0.37	0.60	0.40	0.61
2	0.64	0.11	0.61	0.64	0.64	0.67	0.72	0.75
3	0.53	0.19	0.53	0.53	0.57	0.70	0.57	0.76
4	0.32	0.13	0.25	0.32	0.33	0.44	0.35	0.45
5	0.99	0.49	0.66	0.99	1.00	1.00	0.99	1.00
6	0.62	0.01	0.60	0.62	0.63	0.63	0.62	0.62
7	1.00	0.22	1.00	2.54	1.00	1.00		1.00
8	0.58	0.19	0.49	0.58	0.59	0.81	0.67	0.85
9	0.17	0.05	0.13	0.17	0.19	0.17	0.28	0.41
10	0.78	0.37	0.78	0.78	1.00	0.80	0.78	1.00
11	0.81	0.58	0.81	0.81	1.00	0.95	0.81	1.00
12	0.31	0.13	0.31	0.31	0.31	0.49	0.37	0.50
13	0.32	0.26	0.32	0.32	0.44	0.53	0.32	0.67
14	0.67	0.24	0.58	0.67	0.69	1.00	0.76	1.00
15	0.63	0.26	0.62	0.63	0.63	0.81	0.75	0.92
16	0.88	0.34	0.77	0.88	0.89	1.00	0.90	1.00
17	0.68	0.31	0.68	0.68	0.69	0.82	0.73	0.83
18	0.72	0.27	0.70	0.72	0.72	0.97	0.82	1.00
19	1.00	0.36	1.00	1.47	1.00	1.00	1.00	1.00
20	0.30	0.21	0.30	0.30	0.40	0.35	0.30	0.45
21	1.00	0.10	1.00	1.59	1.00	1.00	1.00	1.00
22	0.30	0.15	0.23	0.30	0.32	0.39	0.31	0.39
23	0.25	0.15	0.23	0.25	0.28	0.37	0.31	0.41
24	0.94	0.63	0.72	0.94	1.00	1.00	0.94	1.00
25	0.69	0.18	0.69	0.69	0.88	0.76	0.69	0.98
26	1.00	1.00	1.00	8.92		1.00	1.00	1.00
27	0.42	0.19	0.42	0.42	0.42	0.65	0.61	0.76

28	0.63	0.09	0.63	0.63	1.00	0.63	0.63	1.00
29	1.00	1.00	1.00	10.00	1.00		1.00	1.00
30	0.59	0.22	0.59	0.59	0.59	0.85	0.63	0.85
31	0.48	0.16	0.46	0.48	0.48	0.68	0.71	0.86
32	0.21	0.07	0.21	0.21	0.21	0.27	0.30	0.36
33	0.54	0.33	0.53	0.54	0.56	0.90	0.61	0.90
34	0.24	0.04	0.24	0.24	0.31	0.24	0.27	0.41
35	1.00	0.15	0.93	1.05	1.00	1.00	1.00	1.00
36	0.33	0.09	0.33	0.33	0.38	0.38	0.48	0.69
37	0.77	0.17	0.77	0.77	0.77	0.84	0.78	0.84
38	0.48	0.15	0.48	0.48	0.59	0.84	0.48	0.97
39	0.44	0.17	0.44	0.44	0.62	0.63	0.49	0.86
40	0.22	0.08	0.19	0.22	0.22	0.29	0.28	0.34

MIN	0.17	0.01	0.13	0.17	0.19	0.17	0.27	0.34
MAX	1.00	1.00	1.00	10.00	1.00	1.00	1.00	1.00
MEAN	0.59	0.25	0.56	1.08	0.63	0.70	0.63	0.79
SD	0.27	0.22	0.26	2.00	0.28	0.26	0.25	0.23
Number efficient units	6.00	2.00	5.00	0.00	10.00	8.00	5.00	14.00

TS 3 “profitability”

	no restrictions	weights imposed	variable removed (Bonus)	superefficiency	DMU removed (DMU 6)
Sensitivity Analysis Category		1	2	3	4
Unit name	Score	Score	Score	Score	Score
1	0.34	0.33	0.34	0.34	0.34
2	0.49	0.44	0.49	0.49	0.49
3	0.31	0.27	0.31	0.31	0.40
4	0.28	0.25	0.28	0.28	0.46
5	1.00	1.00	1.00	1.38	1.00
6	1.00	1.00	1.00	2.23	
7	0.41	0.40	0.41	0.41	0.41
8	0.37	0.36	0.37	0.37	0.38
9	0.20	0.18	0.20	0.20	0.22
10	1.00	1.00	1.00	1.16	1.00
11	0.82	0.71	0.67	0.82	1.00
12	0.74	0.64	0.61	0.74	0.74
13	0.91	0.61	0.91	0.91	0.91
14	0.47	0.45	0.47	0.47	0.47
15	0.70	0.56	0.70	0.70	0.88
16	1.00	1.00	1.00	1.44	1.00
17	1.00	0.89	1.00	1.00	1.00
18	0.37	0.36	0.35	0.37	0.38
19	0.78	0.66	0.78	0.78	0.78
20	0.57	0.54	0.57	0.57	0.57
21	0.83	0.77	0.83	0.83	0.84
22	0.18	0.16	0.13	0.18	0.23

23	0.31	0.25	0.20	0.31	0.34
24	0.93	0.87	0.93	0.93	0.93
25	1.00	0.90	0.87	1.00	1.00
26	0.85	0.72	0.85	0.85	0.85
27	0.31	0.29	0.31	0.31	0.49
28	0.47	0.32	0.47	0.47	0.47
29	0.37	0.32	0.29	0.37	0.41
30	0.36	0.34	0.36	0.36	0.36
31	0.37	0.32	0.37	0.37	0.65
32	0.48	0.38	0.48	0.48	0.48
33	0.27	0.25	0.27	0.27	0.27
34	0.73	0.69	0.73	0.73	0.73
35	0.61	0.57	0.61	0.61	0.61
36	0.56	0.53	0.56	0.56	0.63
37	0.73	0.66	0.73	0.73	0.73
38	0.47	0.42	0.47	0.47	0.86
39	0.68	0.66	0.68	0.68	0.76
40	0.34	0.30	0.34	0.34	0.57

MIN	0.18	0.16	0.13	0.18	0.22
MAX	1.00	1.00	1.00	2.23	1.00
MEAN	0.59	0.53	0.57	0.65	0.63
SD	0.26	0.25	0.26	0.40	0.25
Number efficient units	5.00	0.00	0.00	0.00	6.00

TS 4 “bonuses”

	no restrictions	weights imposed	superefficiency
Sensitivity Analysis Category		1	3
Unit name	Score	Score	Score
1	0.31	0.31	0.31
2	0.49	0.46	0.49
3	0.31	0.30	0.31
4	0.28	0.25	0.28
5	1.00	1.00	1.28
6	1.00	1.00	1.30
7	0.39	0.39	0.39
8	0.35	0.35	0.35
9	0.15	0.15	0.15
10	1.00	1.00	1.16
11	0.82	0.78	0.82
12	0.74	0.70	0.74
13	0.61	0.52	0.61
14	0.47	0.46	0.47
15	0.70	0.62	0.70
16	1.00	1.00	1.44
17	0.74	0.72	0.74
18	0.37	0.37	0.37

19	0.78	0.68	0.78
20	0.57	0.55	0.57
21	0.78	0.78	0.78
22	0.18	0.18	0.18
23	0.31	0.28	0.31
24	0.93	0.90	0.93
25	1.00	0.99	1.00
26	0.85	0.78	0.85
27	0.28	0.28	0.28
28	0.47	0.34	0.47
29	0.37	0.36	0.37
30	0.36	0.35	0.36
31	0.37	0.33	0.37
32	0.33	0.31	0.33
33	0.27	0.26	0.27
34	0.73	0.70	0.73
35	0.61	0.60	0.61
36	0.56	0.53	0.56
37	0.73	0.67	0.73
38	0.47	0.43	0.47
39	0.68	0.66	0.68
40	0.34	0.31	0.34

MIN	0.15	0.15	0.15
MAX	1.00	1.00	1.44
MEAN	0.57	0.54	0.60
SD	0.26	0.26	0.32
Number efficient units	5.00	0.00	0.00

TS 5 “performance targets”

	no restrictions	weights imposed	superefficiency	DMU removed (no. 2)	DMU removed (no. 28)
Sensitivity Analysis Category		1	3	3	3
Unit name	Score	Score	Score	Score	Score
1	0.46	0.43	0.46	0.46	0.46
2	1.00	1.00	2.49		1.00
3	0.55	0.48	0.55	0.55	0.55
4	0.42	0.35	0.42	0.42	0.42
5	1.00	1.00	1.65	1.00	1.00
6	0.73	0.69	0.73	0.73	0.73
7	0.65	0.64	0.65	0.68	0.65
8	0.68	0.58	0.68	0.68	0.68
9	0.58	0.52	0.58	0.66	0.60
10	1.00	1.00	1.24	1.00	1.00
11	0.61	0.60	0.61	0.61	0.64
12	0.79	0.75	0.79	0.79	0.79
13	0.74	0.64	0.74	0.74	0.96
14	0.73	0.68	0.73	0.73	0.73

15	0.90	0.61	0.90	0.90	0.90
16	1.00	1.00	1.16	1.00	1.00
17	0.73	0.69	0.73	0.73	0.73
18	0.56	0.51	0.56	0.56	0.56
19	0.99	0.92	0.99	0.99	0.99
20	0.73	0.71	0.73	0.73	0.80
21	0.70	0.62	0.70	0.70	0.70
22	0.28	0.25	0.28	0.28	0.28
23	0.46	0.35	0.46	0.46	0.46
24	0.96	0.95	0.96	0.96	0.97
25	1.00	1.00	1.45	1.00	1.00
26	1.00	1.00	1.33	1.00	1.00
27	0.34	0.31	0.34	0.34	0.34
28	1.00	1.00	5.27	1.00	
29	0.40	0.34	0.40	0.40	0.40
30	0.66	0.65	0.66	0.85	0.66
31	0.60	0.56	0.60	0.60	0.60
32	0.54	0.45	0.54	0.54	0.67
33	0.32	0.28	0.32	0.32	0.32
34	1.00	1.00	1.13	1.00	1.00
35	0.88	0.84	0.88	0.88	0.96
36	1.00	1.00	1.26	1.00	1.00
37	1.00	1.00	1.10	1.00	1.00
38	0.86	0.78	0.86	0.86	0.86
39	1.00	1.00	1.23	1.00	1.00
40	0.50	0.42	0.50	0.50	0.50

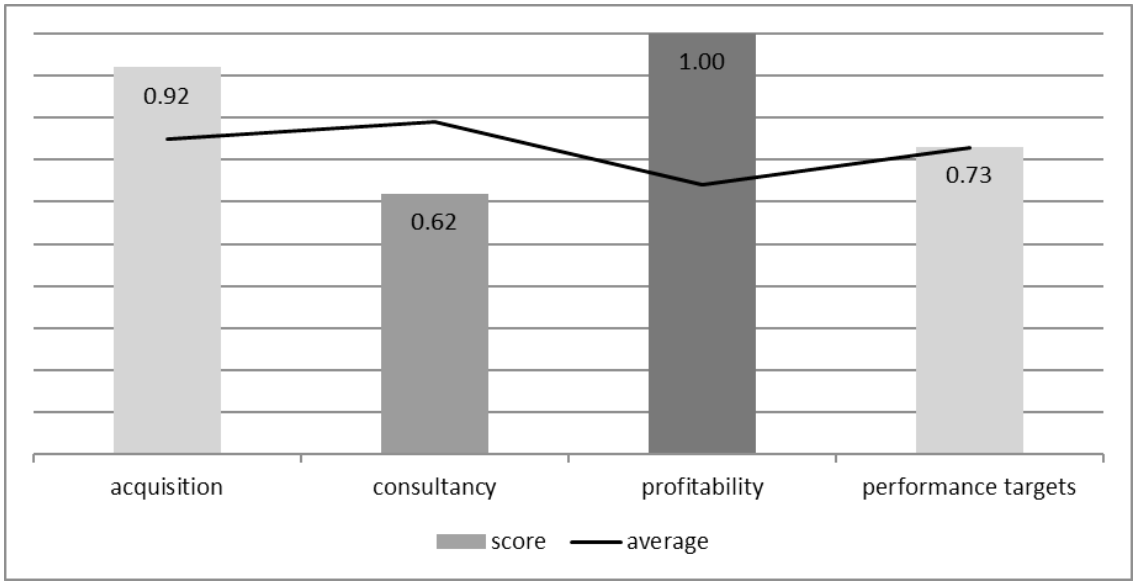
MIN	0.28	0.25	0.28	0.28	0.28
MAX	1.00	1.00	5.27	1.00	1.00
MEAN	0.73	0.69	0.94	0.73	0.74
SD	0.23	0.25	0.82	0.23	0.23
Number efficient units	11.00	11.00	0.00	10.00	10.00

Annex 3: Performance Evaluation Sheet for employee 6

performance evaluation sheet: employee no. 6

stages under evaluation

Overview



Results per stage

acquisition	actual score:	0.92	main peer(s):	20
	average score:	0.75	Reference count:	--

	Measurement criteria	need for improvement	actual	projection
Input variables	CUST		208	
	INHAB		34,209	
	VOLACT		8,559,055 €	
Output variables	APP	○	355	384
	SELL	○	144	189

consultancy and sales	actual score:	0.62	main peer(s):	21, 26, 29
	average score:	0.79	Reference count:	--

	Measurement criteria	Need for improvement	actual	projection
Input variables	APP		355	
	SELL		144	
Output variables	VOLACT	○	10,447,067 €	16,821,230€
	VOLL	●	169,933 €	629,645 €
	BSH	○	205,703 €	331,221 €
	Union	●	3,416 €	89,323 €

profitability	actual score:	1.00	main peer(s):	--
	average score:	0.64	Reference count:	(23)

	Measurement criteria	Need for improvement	actual	projection
Input variables	SAL		51,202 €	
	TRAI		0 €	
	BONUS		0 €	
Output variables	CMII		199,778 €	199,778 €
	CommE		48,076 €	48,076 €

bonuses	actual score:	1.00	main peer(s):	--
	average score:	0.57	Reference count:	18

	Measurement criteria		
Input variables	SAL		
	BONUS	**Actual bonus based on score**	**3,393.43 €**
Output variables	CMII		
	CommE	Average bonus based on score	1,925.00 €

performance targets	actual score:	0.73	main peer(s):	5, 10
	average score:	0.73	Reference count:	

	Measurement criteria	need for improvement	actual	Performa targets
Input variables	CUST		204	
	INHAB		34,209	
	VOLACT		8,559,055 €	
	SAL		51,202 €	
Output variables	CMII	○	199,778 €	274,16
	CommE	○	48,076 €	65,97

TOPIC GUIDE LIST

Focus Group Discussion

- **Topic: Suitability of DEA-based Performance Evaluation System**

I: Opening Questions	- Participant´s introduction - former experience in performance evaluation
II: Introductory Questions	- involvement in case-study (if applicable) - if involved: general impressions/ opinion about the method
III: How does the proposed method meet general requirements?	- are the results understandable? - do managers feel able to explain results to employees? - is the whole procedure understandable? - does the method enable a non-biased performance evaluation? - is a non-biased evaluation even desired by management? - is the method suitable to stimulate continuous improvement/ right behaviour? - is the method suitable to reflect the organization´s business strategy? - is data collection manageable? - does the methodology enable the use of balanced measures? - Are the chosen measures appropriate to capture the employees´ performance?
IV: How does it serve administrative purposes?	- is the method suitable for: • Making promotion decisions • Assigning work more effectively/ making transfer decisions • Making decisions about layoffs and terminations • Determining bonus payments - do the calculated bonuses seem reasonable? - only participants of the case study: did results meet your understanding of the employee´s performances?

V: How does it serve developmental purposes?	- Do the results indicate areas of individual improvement/ Are they suitable to identify training needs? - Does is become evident why some employees receive higher ratings than others? - Are the results suitable for setting performance targets - Are the results a good basis to counsel employees regarding their career goals?
VI: How does it meet the legal requirements of the organisation?	- Are there any legal concerns regarding the implementation of the method?
VII: How does it compare to other evaluation methods?	• **-** How does the method compare to other performance evaluation methods you know/ apply? - What are the pros and cons compared to other methods?
VIII: Ending questions	• - Is the procedure transparent and understandable? - What would be an appropriate frequency to run the evaluation?
IX: Concluding statements	• - what is the perceived cost/benefit relationship? - assessment of general suitability - what was perceived as most important issue during the discussion

Annex 5: Code list

ID	Code
A	**traditional approaches**
A1	previous experience
A2	flaws
A3	strengths
A4	comparison to DEA approach

B	**perceptions of DEA approach**
B1	provides rich data
B2	is fair
B3	provides in-depth results
B4	serves administrative purposes
B5	serves developmental purposes
B6	results confirm impressions of performance
B7	superior to traditional approaches

C	**complexity**
C1	mathematical skills required
C2	major principles need to be understood
C3	performance evaluation sheet reduces complexity

D	**objectivity vs. subjectivity**
D1	provides objective results
D2	reduces bias
D3	objectivity supports perceived fairness
D4	results cannot be influenced
D5	performance evaluation requires subjective components
D6	Performance evaluation should be objective
D7	approach should be supplemented by subjective component

E	**employees´ development**
E1	strengths and weaknesses can be identified
E2	improvement rate is beneficial
E3	improvement rate s.t. too demanding
E4	identification of top-performers
E5	sound basis for counselling

F	**results**
F1	rich and extensive

ID	Code
I	**measures**
I1	to be chosen carefully
I2	not too many
I3	input and output orientation
I4	should be reviewed regularly
I5	not all intended measures are available

J	**data collection**
J1	time consuming
J2	not time consuming if measures are available
J3	need for clear responsibilities

K	**linkage to strategy**
K1	suitable to link strategy to performance
K2	can be linked to BSC
K3	translates strategy to employees
K4	measures need to be chosen right

L	**transferability**
L1	soft skills hard to quantify
L2	criteria for back office employees is not easy to define
L3	criteria for back office employees is available
L4	not suitable for small groups

M	**bonuses**
M1	calculation is fair
M2	calculation is reasonable
M3	not to be paid until a certain score
M4	should be used for all calculations (including management)
M5	other measures for calculating bonuses for management

N	**ratings and raters**
N1	need to gain understanding of procedure
N2	have no influence on results
N3	need training
N4	overwhelmed by data
N5	execution once a year

F2	enable in-depth
F3	too complex
F4	should be made available to employees
F5	restricted access for management
F6	performance evaluation sheet is important
F7	require training for interpretation

O	**peers**
O1	should be named
O2	should not be named
O3	confidential issues
O4	motivational issues
O5	can serve as benchmarks
O6	explain performance gap
O7	need to agree to be named

G	**performance targets**
G1	targets seem reasonable
G2	targets seem too demanding
G3	may not always be achievable
G4	can serve as orientation
G5	are understandable

P	**weights**
P1	should be incorporated
P2	should not be incorporated
P3	reflect management´s priorities
P4	bring bias to the evaluation
P5	reduce flaws of approach

H	**legal requirements**
H1	Remuneration Ordinance
H2	German Equal Treatments act
H3	involvement of worker´s council
H4	data protection

Annex 6: Focus Group protocol (excerpt)

Focus Group ID: A-M
Meeting Date: 09th January 2014 (9:00 – 12:00)

Constituency: Chief executives (SH, BT), board members (MK, MT), head of HR department (IK), head of sales management (AK), branch managers (ML, SK, TP)

Number of Participants: 9

Moderator: SR

NOTES OF DISCUSSION

Question no.	notes	Quotes
I	All participants have been involved in the case study (to a certain degree) and are familiar with the procedure and the results.	•
I	Only two branch managers have little experience with the PES the bank has applied so far. All other participants have broad experience with different methods for measuring employee performance.	IK: *"I was with four organizations before I started working with Bank A. Each organization had its own PES – although the term "system" is excessive in most cases."*
II	The participants all agree that the method is highly objective. Most of the participants evaluated this characteristic as positive.	
I	There is consensus that the proposed method is preferred over the evaluation method the bank currently uses. MT points out that it is by far the most fair evaluation method he has ever experienced. SK emphasized the advantage of being able to make decisions that are not based on pure "gut-feeling".	*TP:" I was surprised that most results met my impression of the employees´ performance."*
II	IK claims that there should be a possibility to consider the manager´s impression of employees. This issue is discussed controversial. ML and TP disagree in favour of objectivity.	*IK: "Although that would make the method vulnerable to bias, a subjective component should be included to get the full picture."* *ML: "A major advantage is that there is no opportunity to influence results."*
II	Several participants emphasize the importance of choosing measures appropriately.	*MK: "If we only had one workshop for defining measures, we would have ended up with more than thirty distinct measures. A second workshop to reduce the number and to see which measures were important and available was crucial."*

Annex 7: Coded Focus Group protocol (excerpt)

Focus Group ID: A-M
Meeting Date: 09th January 2014 (9:00 – 12:00)

Constituency: Chief executives (SH, BT), board members (MK, MT), head of HR department (IK), head of sales management (AK), branch managers (ML, SK, TP)

Number of Participants: 9

Moderator: SR

NOTES OF DISCUSSION_CODED

Question no.	notes	Quotes	Code
I	All participants have been involved in the case study (to a certain degree) and are familiar with the procedure and the results.		A1
I	Only two branch managers have little experience with the PES the bank has applied so far. All other participants have broad experience with different methods for measuring employee performance.	IK: *"I was with four organizations before I started working with Bank A. Each organization had its own PES – although the term "system" is excessive in most cases."*	A1 A1 A2
II	The participants all agree that the method is highly objective. Most of the participants evaluated this characteristic as positive.		D1 B7, A4
I	There is consensus that the proposed method is preferred over the evaluation method the bank currently uses. MT points out that it is by far the most fair evaluation method he has ever experienced. SK emphasized the advantage of being able to make decisions that are not based on pure "gut-feeling".	*TP:" I was surprised that most results met my impression of the employees´ performance."*	B6 B7 B2, B7 D1, D2
II	IK claims that there should be a possibility to consider the manager´s impression of employees. This issue is discussed controversial. ML and TP disagree in favour of objectivity.	*IK: "Although that would make the method vulnerable to bias, a subjective component should be included to get the full picture."* *ML: "A major advantage is that there is no opportunity to influence results."*	D7 B7, D4, D6

Annex 8: Questionnaire (English version)

Evaluation of an approach to evaluate employee performance in the service industry

Please complete the following questionnaire. The survey is conducted anonymously and your responses will be treated confidentially.

Introductory Part

Please assess the requirements listed below considering their importance for a performance evaluation.

	Very important	Quite important	Neither important nor unimportant	Quite unimportant	Very unimportant
I.1 The Performance Evaluation should be fair.	❑	❑	❑	❑	❑
I.2 The measures should be known.	❑	❑	❑	❑	❑
I.3 The Measures should be determined in cooperation with employees	❑	❑	❑	❑	❑
I.4 The measurement should be objective.	❑	❑	❑	❑	❑
I.5 The measures should not only focus on outputs.	❑	❑	❑	❑	❑
I.6 The Performance Evaluation should identify weaknesses and potential for improvement.	❑	❑	❑	❑	❑
I.7The whole process should be comprehensible.	❑	❑	❑	❑	❑
I.8 The results should be used as a basis for calculation bonuses or variable compensation.	❑	❑	❑	❑	❑

Please have a look at the provided results sheet and answer the following questions:

Section A *general requirements*

A.1 The results and statements of the provided result tables are comprehensible.

	Very comprehensible	Quite comprehensible	Neither comprehensible nor incomprehensible	Quite incomprehensible	Very incomprehensible
A.1.1 For measuring individual employee performance	❑	❑	❑	❑	❑
A.1.2 For setting targets	❑	❑	❑	❑	❑
A.1.3 For determining bonuses	❑	❑	❑	❑	❑

A.2 The measures used for measuring performance are comprehensible.

	Very comprehensible	Quite comprehensible	Neither comprehensible nor incomprehensible	Quite incomprehensible	Very incomprehensible
A.2.1 For measuring individual employee performance	❑	❑	❑	❑	❑
A.2.2 For setting targets	❑	❑	❑	❑	❑
A.2.3 For determining bonuses	❑	❑	❑	❑	❑

A.3 The results of the provided result tables are traceable.

	Very traceable	Quite traceable	Neither traceable nor untraceable	Quite untraceable	Very untraceable
A.3.1 For measuring individual employee performance	❑	❑	❑	❑	❑
A.3.2 For setting targets	❑	❑	❑	❑	❑
A.3.3 For determining bonuses	❑	❑	❑	❑	❑

A.4 The PE is fair as a basis for measuring performance.

	Very fair	Quite fair	Neither fair nor unfair	Quite unfair	Very unfair
A.4.1 evaluating individual employee performance	❑	❑	❑	❑	❑
A.4.2 setting targets	❑	❑	❑	❑	❑
A.4.3 determining bonuses	❑	❑	❑	❑	❑

A.5 The PE is an objective basis for measuring performance.

	Very objective	Quite objective	Neither objective nor unobjective	Quite unobjective	Very unobjective
A.5.1. For measuring individual employee performance	❑	❑	❑	❑	❑
A.5.2 For setting targets	❑	❑	❑	❑	❑
A.5.3 For determining bonuses	❑	❑	❑	❑	❑

Section B *administrative and developmental purpose*

B.1 The methodology can help to set ambiguous, but achievable targets.

Strongly agree	agree	Neither agree nor disagree	disagree	Strongly disagree
❑	❑	❑	❑	❑

B.2. The methodology can help to identify individual weaknesses and to improve performance.

Strongly agree	agree	Neither agree nor disagree	disagree	Strongly disagree
❑	❑	❑	❑	❑

B.3. The methodology can help to identify individual strengths and to develop further.

Strongly agree	agree	Neither agree nor disagree	disagree	Strongly disagree
❑	❑	❑	❑	❑

B.4. The methodology can help to recognize where a participant is better than others.

Strongly agree	agree	Neither agree nor disagree	disagree	Strongly disagree
❑	❑	❑	❑	❑

Section C *Comparison to other approaches*

C.1. How does the methodology compare to other approaches or procedures for evaluating employee performance you know?

	better	worse	About the same	I do not know the methodology/procedure
C.1.1 Simple rankings	❑	❑	❑	❑
C.1.2 Written appraisal (by management)	❑	❑	❑	❑
C1.3 Appraisal interview	❑	❑	❑	❑
C1.4 360 degree feedback	❑	❑	❑	❑
C.1.5 Other methods (please note)	❑	❑	❑	❑

C.2. How does the methodology compare to other approaches or procedures for determining performance targets?

	better	worse	About the same	I do not know the methodology/procedure
C.2.1 last year´s targets + x	❑	❑	❑	❑
C.2.2 Same targets for everyone	❑	❑	❑	❑
C.2.3 Other methods (please note)	❑	❑	❑	❑

C.3. How does the methodology compare to other approaches for determining bonuses or other incentives?

	better	worse	About the same	I do not know the methodology/procedure
C.3.1 Decision by management	❑	❑	❑	❑
C.3.2. Fixed share of corporate profit	❑	❑	❑	❑
C.3.3 Other methods (please note)	❑	❑	❑	❑

Section D *Overall assessment*

D.1. Please asses the general suitability of the proposed approach for the following purposes:

	Very suitable	suitable	Neither suitable nor unsuitable	unsuitable	Very unsuitable
D.1.1 For evaluating individual employee performance	❑	❑	❑	❑	❑
D.1.2 For determining targets	❑	❑	❑	❑	❑
D.1.3 For determining bonuses	❑	❑	❑	❑	❑

D.2. Would you endorse to apply the methodology in your organization?

	Yes, absolutely	Yes, but complemented by other procedures	Yes, but equally with other procedures	Yes, but only as a supplement to other procedures	no
D.2.1 For evaluating individual employee performance	❑	❑	❑	❑	❑
D.2.2 For determining targets	❑	❑	❑	❑	❑
D.2.3 For determining bonuses	❑	❑	❑	❑	❑

Section E *biographical data*

E.1. Gender

❑ female ❑ male

E.2. Age

under 25	25 - 35	36 - 45	46 - 60	over 60
❑	❑	❑	❑	❑

E.3. Work experience

Less than 1 year	1-5 years	6-10 years	11-20 years	More than 20 years
❑	❑	❑	❑	❑

Frequency Tables Introductory Part

I.1 The PE should be fair

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very important	112	91.8	91.8	91.8
	important	10	8.2	8.2	100.0
	Total	122	100.0	100.0	

I.2 The Measures should be known

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very important	54	44.3	44.3	44.3
	important	46	37.7	37.7	82.0
	neither important nor unimportant	16	13.1	13.1	95.1
	unimportant	4	3.3	3.3	98.4
	very unimportant	2	1.6	1.6	100.0
	Total	122	100.0	100.0	

I.3 The Measures should be determined in cooperation with employees

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very important	8	6.6	6.6	6.6
	important	40	32.8	32.8	39.3
	neither important nor unimportant	40	32.8	32.8	72.1
	unimportant	20	16.4	16.4	88.5
	very unimportant	14	11.5	11.5	100.0
	Total	122	100.0	100.0	

I.4 The Method should be objective

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very important	78	63.9	63.9	63.9
	important	44	36.1	36.1	100.0
	Total	122	100.0	100.0	

I.5 The Measurement should not only focus on outputs

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very important	68	55.7	55.7	55.7
	important	40	32.8	32.8	88.5
	neither important nor unimportant	8	6.6	6.6	95.1
	unimportant	4	3.3	3.3	98.4
	very unimportant	2	1.6	1.6	100.0
	Total	122	100.0	100.0	

I.6 The PE should identify weaknesses and potential for improvement

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very important	76	62.3	62.3	62.3
	important	46	37.7	37.7	100.0
	Total	122	100.0	100.0	

I.7 The whole PE process should be comprehensible

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very important	74	60.7	60.7	60.7
	important	44	36.1	36.1	96.7
	neither important nor unimportant	2	1.6	1.6	98.4
	unimportant	2	1.6	1.6	100.0
	Total	122	100.0	100.0	

I.8 The results should be used as a basis for calculation bonuses or variable compensation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very important	18	14.8	15.0	15.0
	important	70	57.4	58.3	73.3
	neither important nor unimportant	32	26.2	26.7	100.0
	Total	120	98.4	100.0	
Missing	System	2	1.6		
Total		122	100.0		

Frequency Tables Part A

A.1.1 The results and statements of the provided result tables are comprehensible (for measuring performance)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very comprehensible	26	21.3	21.3	21.3
	compehensible	76	62.3	62.3	83.6
	neither comprehenisble nor incomprehensible	12	9.8	9.8	93.4
	incomprehensible	8	6.6	6.6	100.0
	Total	122	100.0	100.0	

A.1.2 The results and statements of the provided result tables are comprehensible (for setting targets)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very comprehensible	16	13.1	13.1	13.1
	compehensible	84	68.9	68.9	82.0
	neither comprehenisble nor incomprehensible	6	4.9	4.9	86.9
	incomprehensible	12	9.8	9.8	96.7
	very incomprehensible	4	3.3	3.3	100.0
	Total	122	100.0	100.0	

A.1.3 The results and statements of the provided result tables are comprehensible (for calculating bonuses)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very comprehensible	28	23.0	23.0	23.0
	compehensible	62	50.8	50.8	73.8
	neither comprehenisble nor incomprehensible	14	11.5	11.5	85.2
	incomprehensible	16	13.1	13.1	98.4
	very incomprehensible	2	1.6	1.6	100.0
	Total	122	100.0	100.0	

A.2.1 The measures used for measuring performance are comprehensible

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very comprehensible	22	18.0	18.0	18.0
	compehensible	74	60.7	60.7	78.7
	neither comprehenisble nor incomprehensible	18	14.8	14.8	93.4
	incomprehensible	8	6.6	6.6	100.0
	Total	122	100.0	100.0	

A.2.2 The measures used for setting targets are comprehensible

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very comprehensible	20	16.4	16.4	16.4
	compehensible	80	65.6	65.6	82.0
	neither comprehenisble nor incomprehensible	12	9.8	9.8	91.8
	incomprehensible	10	8.2	8.2	100.0
	Total	122	100.0	100.0	

A.2.3 The measures used for calculating bonuses are comprehensible

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very comprehensible	24	19.7	19.7	19.7
	compehensible	58	47.5	47.5	67.2
	neither comprehenisble nor incomprehensible	22	18.0	18.0	85.2
	incomprehensible	18	14.8	14.8	100.0
	Total	122	100.0	100.0	

A.3.1 The results of the provided result tables are traceable (for measuring performance)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very traceable	40	32.8	32.8	32.8
	traceable	64	52.5	52.5	85.2
	neither traceable nor untraceable	2	1.6	1.6	86.9
	untraceable	16	13.1	13.1	100.0
	Total	122	100.0	100.0	

A.3.2 The results of the provided result tables are traceable (for setting targets)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very traceable	30	24.6	24.6	24.6
	traceable	72	59.0	59.0	83.6
	neither traceable nor untraceable	12	9.8	9.8	93.4
	untraceable	8	6.6	6.6	100.0
	Total	122	100.0	100.0	

A.3.3 The results of the provided result tables are traceable (for calculating bonuses)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very traceable	36	29.5	29.5	29.5
	traceable	56	45.9	45.9	75.4
	neither traceable nor untraceable	20	16.4	16.4	91.8
	untraceable	10	8.2	8.2	100.0
	Total	122	100.0	100.0	

A.4.1 The PE is fair as a basis for measuring performance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very fair	38	31.1	31.1	31.1
	fair	60	49.2	49.2	80.3
	neither fair nor unfair	18	14.8	14.8	95.1
	unfair	6	4.9	4.9	100.0
	Total	122	100.0	100.0	

A.4.2 The PE is fair as a basis for setting performance targets

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very fair	26	21.3	21.3	21.3
	fair	74	60.7	60.7	82.0
	neither fair nor unfair	16	13.1	13.1	95.1
	unfair	6	4.9	4.9	100.0
	Total	122	100.0	100.0	

A.4.3 The PE is fair as a basis for calculating bonuses

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very fair	24	19.7	19.7	19.7
	fair	60	49.2	49.2	68.9
	neither fair nor unfair	30	24.6	24.6	93.4
	unfair	8	6.6	6.6	100.0
	Total	122	100.0	100.0	

A.5.1 The PE is an objective basis for measuring performance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very objective	20	16.4	16.4	16.4
	objective	88	72.1	72.1	88.5
	neither objective nor unobjective	10	8.2	8.2	96.7
	unobjective	4	3.3	3.3	100.0
	Total	122	100.0	100.0	

A.5.2 The PE is an objective basis for setting performance targets

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very objective	22	18.0	18.0	18.0
	objective	78	63.9	63.9	82.0
	neither objective nor unobjective	18	14.8	14.8	96.7
	unobjective	4	3.3	3.3	100.0
	Total	122	100.0	100.0	

A.5.3 The PE is an objective basis for calculating bonuses

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very objective	22	18.0	18.0	18.0
	objective	66	54.1	54.1	72.1
	neither objective nor unobjective	28	23.0	23.0	95.1
	unobjective	6	4.9	4.9	100.0
	Total	122	100.0	100.0	

Frequency Tables Part B

B.1 The PE can help to set ambitious, but achievable targets

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	34	27.9	27.9	27.9
	agree	62	50.8	50.8	78.7
	neither agree nor disagree	22	18.0	18.0	96.7
	disagree	4	3.3	3.3	100.0
	Total	122	100.0	100.0	

B.2 The PE can help to identify individual weaknesses and to improve performance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	38	31.1	31.1	31.1
	agree	68	55.7	55.7	86.9
	neither agree nor disagree	12	9.8	9.8	96.7
	disagree	4	3.3	3.3	100.0
	Total	122	100.0	100.0	

B.3 The PE can help to identify individual strengths and to develop them further

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	46	37.7	37.7	37.7
	agree	68	55.7	55.7	93.4
	neither agree nor disagree	6	4.9	4.9	98.4
	disagree	2	1.6	1.6	100.0
	Total	122	100.0	100.0	

B.4 The PE can help to recognize where a participant is better than others

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	strongly agree	48	39.3	39.3	39.3
	agree	52	42.6	42.6	82.0
	neither agree nor disagree	14	11.5	11.5	93.4
	disagree	8	6.6	6.6	100.0
	Total	122	100.0	100.0	

Frequency Tables Part C

C.1.1 How does the methodology compare to simple rankings?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	better	98	80.3	80.3	80.3
	worse	8	6.6	6.6	86.9
	about the same	10	8.2	8.2	95.1
	I do not know	6	4.9	4.9	100.0
	Total	122	100.0	100.0	

C.1.2 How does the methodology compare to a written evaluation by management?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	better	78	63.9	65.0	65.0
	worse	16	13.1	13.3	78.3
	about the same	22	18.0	18.3	96.7
	I do not know	4	3.3	3.3	100.0
	Total	120	98.4	100.0	
Missing	System	2	1.6		
Total		122	100.0		

C.1.3 How does the methodology compare to a an appraisal interview?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	better	72	59.0	60.0	60.0
	worse	14	11.5	11.7	71.7
	about the same	30	24.6	25.0	96.7
	I do not know	4	3.3	3.3	100.0
	Total	120	98.4	100.0	
Missing	System	2	1.6		
Total		122	100.0		

C.1.4 How does the methodology compare to a 360 degree feedback?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	better	22	18.0	22.0	22.0
	worse	4	3.3	4.0	26.0
	about the same	24	19.7	24.0	50.0
	I do not know	50	40.9	48.0	100.0
	Total	100	82.0	100.0	
Missing	System	22	18.0		
Total		122	100.0		

C.1.5 How does the methodology compare to any other PA methods you know?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	better	6	4.9	42.9	42.9
	I do not know	6	4.9	42.9	85.7
	about the same	2	1.6	14.3	100.0
	Total	14	11.5	100.0	
Missing	System	108	88.5		
Total		122	100.0		

C.2.1 How does the methodology compare to other methodologies or procedures for target setting: last year´s targets + x

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	better	92	75.4	78.0	78.0
	worse	6	4.9	5.1	83.1
	about the same	16	13.1	13.6	96.6
	I do not know	4	3.3	3.4	100.0
	Total	118	96.7	100.0	
Missing	System	4	3.3		
Total		122	100.0		

C.2.2 How does the methodology compare to other methodologies or procedures for target setting: same targets for everyone

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	better	80	65.6	71.4	71.4
	worse	8	6.6	7.1	78.6
	about the same	16	13.1	14.3	92.9
	I do not know	8	6.6	7.1	100.0
	Total	112	91.8	100.0	
Missing	System	10	8.2		
Total		122	100.0		

C.2.3 How does the methodology compare to other methodologies or procedures for target setting: any other

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	better	8	6.6	44.4	44.4
	worse	4	3.3	22.2	66.7
	about the same	2	1.6	11.1	77.8
	I do not know	4	3.3	22.2	100.0
	Total	18	14.8	100.0	
Missing	System	104	85.2		
Total		122	100.0		

C.3.1 How does the methodology compare to other methodologies or procedures for calculating bonuses: decision by management

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	better	96	78.7	80.0	80.0
	worse	8	6.6	6.7	86.7
	about the same	16	13.1	13.3	100.0
	Total	120	98.4	100.0	
Missing	System	2	1.6		
Total		122	100.0		

C.3.2 How does the methodology compare to other methodologies or procedures for calculating bonuses:fixed share of corporate profit

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	better	88	72.1	75.9	75.9
	worse	14	11.5	12.1	87.9
	about the same	14	11.5	12.1	100.0
	Total	116	95.1	100.0	
Missing	System	6	4.9		
Total		122	100.0		

C.3.3 How does the methodology compare to other methodologies or procedures for calculating bonuses: any other

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	better	6	4.9	60.0	60.0
	I do not know	4	3.3	40.0	100.0
	Total	10	8.2	100.0	
Missing	System	112	91.8		
Total		122	100.0		

Frequency Tables Part D

D.1.1 General suitability of the methodology for the purpose of measuring performance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very suitable	44	36.1	36.1	36.1
	suitable	64	52.5	52.5	88.5
	neither suitable nor unsuitable	8	6.6	6.6	95.1
	unsuitable	6	4.9	4.9	100.0
	Total	122	100.0	100.0	

D.1.2 General suitability of the methodology for the purpose of setting targets

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very suitable	40	32.8	32.8	32.8
	suitable	68	55.7	55.7	88.5
	neither suitable nor unsuitable	8	6.6	6.6	95.1
	unsuitable	6	4.9	4.9	100.0
	Total	122	100.0	100.0	

D.1.3 General suitability of the methodology for the purpose of calculating bonuses

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	very suitable	36	29.5	29.5	29.5
	suitable	60	49.2	49.2	78.7
	neither suitable nor unsuitable	16	13.1	13.1	91.8
	unsuitable	10	8.2	8.2	100.0
	Total	122	100.0	100.0	

D.2.1 Would you endorse to apply the methodology in your company for measuring performance

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, absolutely	26	21.3	22.0	22.0
	yes, but complemented by other procedures	54	44.3	45.8	67.8
	yes, but equally with other procedures	26	21.3	22.0	89.8
	yes, as a supplement to other procedures	4	3.3	3.4	93.2
	no	8	6.6	6.8	100.0
	Total	118	96.7	100.0	
Missing	System	4	3.3		
Total		122	100.0		

D.2.2 Would you endorse to apply the methodology in your company for setting targets

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, absolutely	26	21.3	21.3	21.3
	yes, but complemented by other procedures	52	42.6	42.6	63.9
	yes, but equally with other procedures	24	19.7	19.7	83.6
	yes, as a supplement to other procedures	8	6.6	6.6	90.2
	no	12	9.8	9.8	100.0
	Total	122	100.0	100.0	

D.2.3 Would you endorse to apply the methodology in your company for calculating bonuses

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes, absolutely	28	23.0	23.3	23.3
	yes, but complemented by other procedures	46	37.7	38.3	61.7
	yes, but equally with other procedures	20	16.4	16.7	78.3
	yes, as a supplement to other procedures	12	9.8	10.0	88.3
	no	14	11.5	11.7	100.0
	Total	120	98.4	100.0	
Missing	System	2	1.6		
Total		122	100.0		

Frequency Tables Part E

E.1 Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	male	52	42.6	42.6	42.6
	female	70	57.4	57.4	100.0
	Total	122	100.0	100.0	

E.2 Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	under 25	4	3.3	3.3	3.3
	25 - 35	62	50.8	50.8	54.1
	36 - 45	36	29.5	29.5	83.6
	46 - 60	20	16.4	16.4	100.0
	Total	122	100.0	100.0	

E.3 Work experience

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	less than 1 year	2	1.6	1.6	1.6
	1 - 5 years	18	14.8	14.8	16.4
	6 - 10 years	36	29.5	29.5	45.9
	11 - 20 years	40	32.8	32.8	78.7
	more than 20 years	26	21.3	21.3	100.0
	Total	122	100.0	100.0	

E.4 Previous Experience with Performance Evaluation methods

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	90	73.8	73.8	73.8
	no	32	26.2	26.2	100.0
	Total	122	100.0	100.0	

E.5 Previous Experience with Performance Targets

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	80	65.6	65.6	65.6
	no	42	34.4	34.4	100.0
	Total	122	100.0	100.0	

E.6 Previous Experience with Bonuses

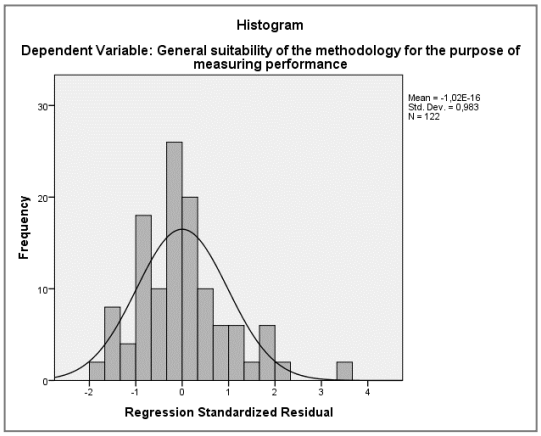
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	yes	62	50.8	50.8	50.8
	no	60	49.2	49.2	100.0
	Total	122	100.0	100.0	

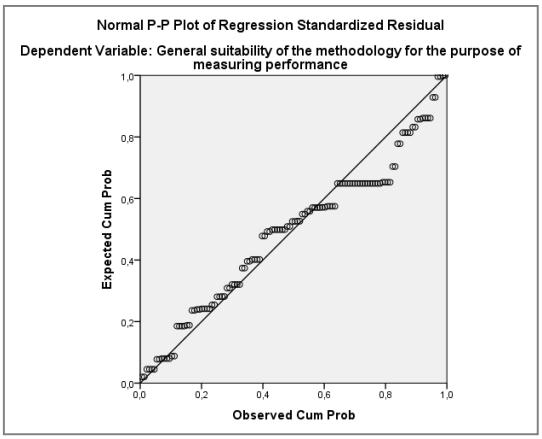
Testing Assumptions number 3 (no or little multicollinearity) and number 4 (independence of observations (no auto correlation))

Model		Collinearity Statistics		auto-correlation
		Tolerance	VIF	Durbin-Watson
A: methods ability to meet requirements for measuring performance * overall suitability for measuring performance	The results and statements of the provided result tables are comprehensible (for measuring performance)	0.364	2.749	1.518
	The measures used for measuring performance are comprehensible	0.466	2.146	
	The results of the provided result tables are traceable (for measuring performance)	0.502	1.993	
	The PE is fair as a basis for measuring performance	0.518	1.930	
	The PE is an objective basis for measuring performance	0.608	1.646	
B: methods ability to meet requirements for determining targets * overall suitability for determining targets	The results and statements of the provided result tables are comprehensible (for setting targets)	0.396	2.527	1.851
	The measures used for setting targets are comprehensible	0.508	1.968	
	The results of the provided result tables are traceable (for setting targets)	0.436	2.295	
	The PE is fair as a basis for setting performance targets	0.554	1.806	
	The PE is an objective basis for setting performance targets	0.550	1.820	
C: methods ability to meet requirements for determining bonuses * overall suitability for determining bonuses	The results and statements of the provided result tables are comprehensible (for calculating bonuses)	0.287	3.490	1.868
	The measures used for calculating bonuses are comprehensible	0.402	2.486	
	The results of the provided result tables are traceable (for calculating bonuses)	0.310	3.229	
	The PE is fair as a basis for calculating bonuses	0.466	2.148	
	The PE is an objective basis for calculating bonuses	0.529	1.890	
D: administrative and developmental purposes * overall suitability for measuring	The PE can help to set ambitious, but achievable targets	0.643	1.555	1.986
	The PE can help to identify individual weaknesses	0.489	2.046	
	The PE can help to identify individual strengths and to develop them further	0.598	1.673	
	The PE can help to recognize where a participant is better than others	0.792	1.263	

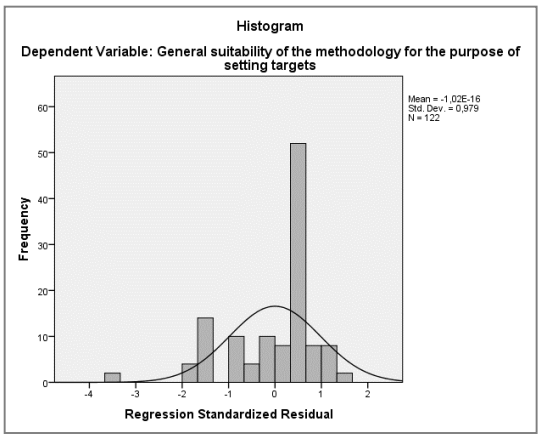
Testing Assumptions number 5 (Multivariate normality) and number 6 (Homoscedasticity)

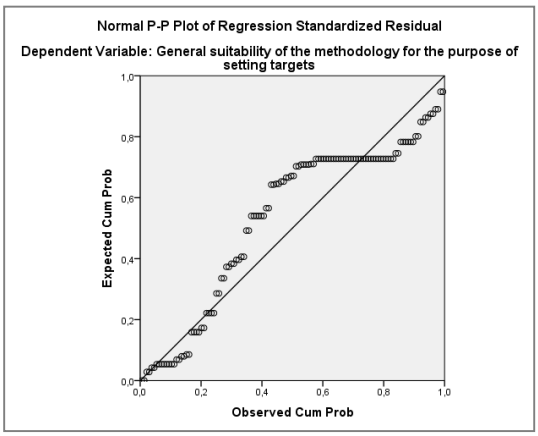
Model A



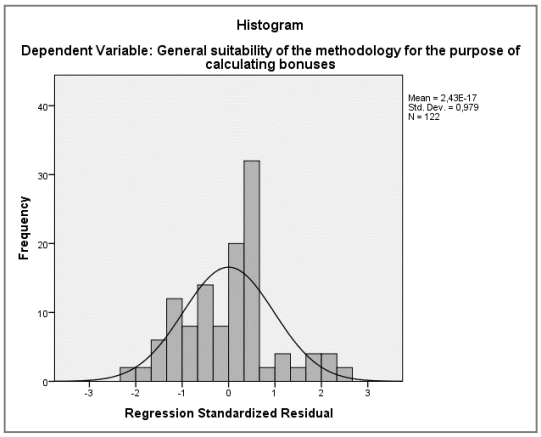


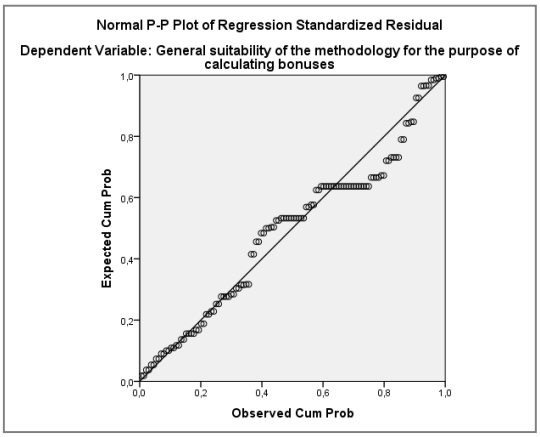
Model B



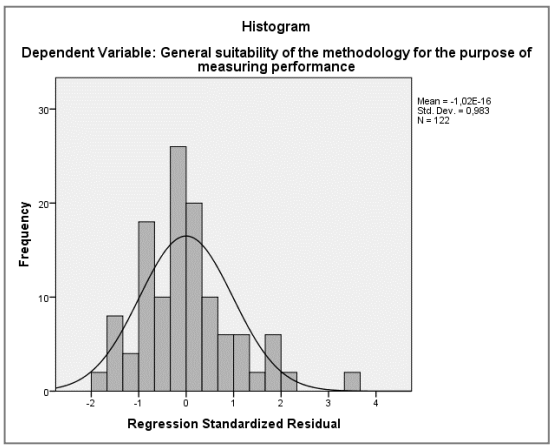


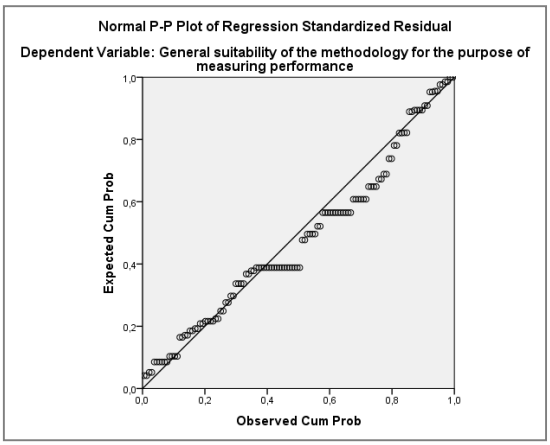
Model C





Model D


Histogram
Dependent Variable: General suitability of the methodology for the purpose of measuring performance
Mean = -1,02E-16
Std. Dev. = 0,983
N = 122
Frequency
Regression Standardized Residual


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: General suitability of the methodology for the purpose of measuring performance
Expected Cum Prob
Observed Cum Prob

Model A: general requirements * general suitability for measuring performance

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.701[a]	0.491	0.469	0.559	1.518

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	34.995	5	6.999	22.376	,000[b]
	Residual	36.284	116	0.313		
	Total	71.279	121			

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	0.218	0.194		1.125	0.263
	The results and statements of the provided result tables are comprehensible (for measuring performance)	0.131	0.111	0.130	1.185	0.238
	The measures used for measuring performance are comprehensible	0.361	0.097	0.360	3.713	0.000
	The results of the provided result tables are traceable (for measuring performance)	-0.109	0.077	-0.132	-1.413	0.160
	The PE is fair as a basis for measuring performance	0.394	0.087	0.416	4.519	0.000
	The PE is an objective basis for measuring performance	0.006	0.106	0.005	0.056	0.955

Model B: general requirements * general suitability for determining performance targets

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.810[a]	0.656	0.641	0.451	1.851

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	45.077	5	9.015	44.229	,000[b]
	Residual	23.645	116	0.204		
	Total	68.721	121			

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	0.182	0.148		1.231	0.221
	The results and statements of the provided result tables are comprehensible (for setting targets)	0.526	0.072	0.636	7.345	0.000
	The measures used for setting targets are comprehensible	-0.109	0.075	-0.110	-1.446	0.151
	The results of the provided result tables are traceable (for setting targets)	-0.025	0.080	-0.026	-0.316	0.753
	The PE is fair as a basis for setting performance targets	0.301	0.075	0.295	4.025	0.000
	The PE is an objective basis for setting performance targets	0.080	0.081	0.072	0.979	0.330

a. Dependent Variable: General suitability of the methodology for the purpose of setting targets

Model C: general requirements * general suitability for calculating bonuses

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.819[a]	0.671	0.657	0.511	1.868

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	61.715	5	12.343	47.278	.000[b]
	Residual	30.285	116	0.261		
	Total	92.000	121			

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	0.095	0.151		0.625	0.533
	The results and statements of the provided result tables are comprehensible (for calculating bonuses)	0.335	0.087	0.382	3.837	0.000
	The measures used for calculating bonuses are comprehensible	0.089	0.077	0.097	1.156	0.250
	The results of the provided result tables are traceable (for calculating bonuses)	-0.040	0.094	-0.041	-0.428	0.669
	The PE is fair as a basis for calculating bonuses	0.559	0.083	0.528	6.761	0.000
	The PE is an objective basis for calculating bonuses	-0.080	0.083	-0.071	-0.963	0.337

a. Dependent Variable: General suitability of the methodology for the purpose of calculating bonuses

Model D administrative and developmental purpose * general suitability for measuring performance

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.550[a]	0.302	0.278	0.652	1.986

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	21.536	4	5.384	12.664	.000[b]
	Residual	49.743	117	0.425		
	Total	71.279	121			

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	0.477	0.203		2.351	0.020
	The PE can help to set ambitious, but achievable targets	0.309	0.096	0.311	3.225	0.002
	The PE can help to identify individual weaknesses and to improve performance	0.185	0.117	0.174	1.577	0.118
	The PE can help to identify individual strengths and to develop them further	0.142	0.120	0.118	1.183	0.239
	The PE can help to recognize where a participant is better than others	0.072	0.077	0.082	0.940	0.349

a. Dependent Variable: General suitability of the methodology for the purpose of measuring performance

Results for Chi Square Analysis

dependent variable: General suitability of the methodology (for measuring employee performance)				
independent variable	Phi	Cramer´s V	p-value (sign.)	p-value (Fishers exact)
Age	.538	.310	.000	.000
Experience with Performance Evaluation	.409	.409	.000	.000
Gender	.388	.388	.000	.000
Work experience	.435	.251	.027	.049

dependent variable: Recommendation of the methodology (for measuring employee performance)				
independent variable	Phi	Cramer´s V	p-value (sign.)	p-value (Fishers exact)
Age	.472	.272	.010	.032
Experience with Performance Evaluation	.238	.238	.153	.173
Gender	.124	.124	.770	.769
Work experience	.461	.231	.067	.057

dependent variable: General suitability of the methodology (for determining performance targets)				
independent variable	Phi	Cramer´s V	p-value (sign.)	p-value (Fishers exact)
Age	.551	.318	.000	.000
Experience with Performance Targets	.252	.252	.046	.039
Gender	.241	.241	.064	0.51
Work experience	.319	.184	.411	.267

dependent variable: Recommendation of the methodology (for determining performance targets)				
independent variable	Phi	Cramer´s V	p-value (sign.)	p-value (Fishers exact)
Age	.507	.293	.000	.000
Experience with Performance Targets	.163	.103	.517	.574
Gender	.272	.272	.058	.057
Work experience	.412	.206	.186	.244

dependent variable: General suitability of the methodology (for determining bonuses)				
independent variable	Phi	Cramer´s V	p-value (sign.)	p-value (Fishers exact)
Age	.554	.320	.000	.000
Experience with Bonuses	.154	.154	.411	.425
Gender	.263	.263	.036	.031
Work experience	.361	.209	.418	.192

dependent variable: Recommendation of the methodology (for determining bonuses)				
independent variable	Phi	Cramer´s V	p-value (sign.)	p-value (Fishers exact)
Age	.557	.321	.000	.000
Experience with Bonuses	.139	.139	.679	.414
Gender	.221	.221	.214	.224
Work experience	.481	.241	.032	.066